510

THE QUARTERLY JOURNAL OF MATHEMATICS

OXFORD SECOND SERIES

Volume 3 *No.* 12 *December* 1952

H. Schneider: Theorems on Normal Matrices . . . 241

A. J. A. Morgan: The Reduction by One of the Number of Independent Variables in some Systems of Partial Differential Equations 250

C. R. Putnam: Function Space: Hilbert-space Correspondences in Quantum Mechanics . . . 260

A. M. Macbeath: Non-homogenous Lattices in the Plane . 268

A. Y. Fawaz: On an Unsolved Problem in the Analytic Theory of Numbers 282

H. G. Eggleston: A Proof of Blaschke's Theorem on the Reuleaux Triangle 296

A. H. Stone: On Infinitely Multicoherent Spaces . . 298

D. D. Kosambi: Path Geometry and Continuous Groups . 307

THE JOHN CRERAR LIBRARY

OXFORD
AT THE CLARENDON PRESS

Price 12*s.* 6*d. net*

PRINTED IN GREAT BRITAIN BY CHARLES BATEY AT THE UNIVERSITY PRESS, OXFORD

THE QUARTERLY JOURNAL OF
MATHEMATICS

OXFORD SECOND SERIES

Edited by T. W. CHAUNDY, U. S. HASLAM-JONES,
J. H. C. THOMPSON

THE QUARTERLY JOURNAL OF MATHEMATICS (OXFORD SECOND SERIES) is published at 12s. 6d. net for a single number with an annual subscription (for four numbers) of 44s. 6d. post free.

Papers, of a length normally not exceeding 20 printed pages of the Journal, are invited on subjects of Pure and Applied Mathematics, and should be addressed 'The Editors, Quarterly Journal of Mathematics, Clarendon Press, Oxford'. They should be in a condition fit for sending to a referee and in form and notation convenient for printing. The Editors as a rule will not wish to accept material that they cannot see their way to publish within a twelvemonth.

While every care is taken of manuscripts submitted for publication, the Publisher and the Editors cannot hold themselves responsible for any loss or damage. Authors are advised to retain a copy of anything they may send for publication. Authors of papers printed in the Quarterly Journal will be entitled to 50 free offprints.

Correspondence on the *subject-matter* of the Quarterly Journal should be addressed, as above, to 'The Editors', at the Clarendon Press. All other correspondence should be addressed to the Publisher:—

GEOFFREY CUMBERLEGE
OXFORD UNIVERSITY PRESS
AMEN HOUSE, LONDON, E.C.4

The publishers are signatories to the Fair Copying Declaration in respect of this journal. Details of the Declaration may be obtained from the offices of the Royal Society upon application.

THEOREMS ON NORMAL MATRICES

By H. SCHNEIDER (*Edinburgh*)

[Received 12 September 1951; in revised form 10 December 1951]

1. As a corollary to general considerations on matrices A, B whose commutator $AB-BA$ is nilpotent, Drazin, Dungey, and Gruenberg (2) have recently obtained this theorem on normal matrices:† '*Let A and B be normal matrices whose latent roots are α_i, β_i $(i = 1,..., n)$. Then A and B commute if and only if the latent roots of every scalar polynomial $f(A, B)$ are $f(\alpha_i, \beta_i)$ $(i = 1,..., n)$ for some ordering α_i, β_i.*'

Using a different approach I shall show that for the commutativity of the normal matrices A and B it is sufficient to make apparently weaker assumptions. In these $f(A, B)$ will be considerably specialized and a condition, implied by the theorem quoted, will be imposed on the sum of squared moduli of latent roots only.

Every square matrix A has a 'polar representation' $A = HU$, where H is non-negative definite Hermitian, U is unitary [Wintner and Murnaghan (7), Williamson (6)]. The matrices H and U commute if and only if A is normal. If A, B, and AB are normal, it is known that the matrices of the polar representation of A and B commute in pairs, except possibly the unitary pair, provided that the latter are properly chosen.‡ This suggested that, if polar representations were considered, a similar theorem might be obtained for normal, but not necessarily commutative, matrices, whose product is also normal.

To avoid repetition in the statement of the theorems I shall assume throughout that A and B are normal $n \times n$ matrices with complex elements and latent roots α_i, β_i $(i = 1,..., n)$ respectively. Also that

$$A = HU = UH \quad \text{and} \quad B = KV = VK$$

are the polar representations of A and B, where H and K are non-negative definite Hermitian, U and V unitary.

† A review of several important properties of normal matrices may be found in Drazin (1). I shall generally use the results of that paper without giving further references.

‡ Wiegmann (5). The unitary polar matrix of a singular matrix is not unique. The proviso is necessary when the rank of either A or B is less than or equal to $n-2$.

Quart. J. Math. Oxford (2), 3 (1952), 241-9.

2. My results rest on a property of matrices due to Schur (**4**):
Let C be a matrix with latent roots γ_i $(i = 1,..., n)$; then

$$\sum_{i=1}^{n} |\gamma_i|^2 \leqslant \sum_{i,j=1}^{n} |c_{ij}|^2,$$

and

$$\sum_{i=1}^{n} |\gamma_i|^2 = \sum_{i,j=1}^{n} |c_{ij}|^2$$

if and only if C is normal.

I shall first prove a lemma.

LEMMA. *Let C be a normal $n \times n$ matrix, whose latent roots are γ_{ij} $(j = 1,..., n_i;\ i = 1,..., k)$, where $\sum_{i=1}^{k} n_i = n$. Let C be partitioned so that the k matrices C_{ii} in the diagonal are square. If the sum of squared moduli of latent roots of C_{ii} is $\sum_{j=1}^{n_i} |\gamma_{ij}|^2$, for all i, then C is the direct sum*

$$C = C_{11} \dot{+} C_{22} \dot{+} \ldots \dot{+} C_{kk}.$$

Proof. Let S_i be the sum of squared moduli of elements of C_{ii} and let

$$S = \sum_{i,j=1}^{n} |c_{ij}|^2.$$

Using Schur's theorem for C_{ii} we have

$$\sum_{j=1}^{n} |\gamma_{ij}|^2 \leqslant S_i \quad (i = 1,..., k).$$

Since C is normal,

$$S = \sum_{i=1}^{k} \sum_{j=1}^{n_i} |\gamma_{ij}|^2 \leqslant \sum_{i=1}^{k} S_i.$$

But $S > \sum_{i=1}^{k} S_i$; unless $C_{ij} = 0$ $(i \neq j)$, when $S = \sum_{i=1}^{k} S_i$. It follows that $C_{ij} = 0$ $(i \neq j)$ and therefore that

$$C = C_{11} \dot{+} C_{22} \dot{+} \ldots \dot{+} C_{kk}.$$

It will be convenient to denote the sum of squared moduli of latent roots of a matrix A by $\Lambda(A)$.

THEOREM 1. *The matrices A and B commute if and only if the sum of squared moduli of latent roots of every scalar polynomial $f(A)B$ is*

$$\sum_{i=1}^{n} |f(\alpha_i)\beta_i|^2$$

for some ordering α_i, β_i $(i = 1,..., n)$.

Proof. By a well-known theorem of Frobenius there is an ordering for which the latent roots of every $f(A)B$ are $f(\alpha_i)\beta_i$, provided that A and B commute. This ordering obviously satisfies the theorem.

Now suppose there is an ordering for which

$$\Lambda\{f(A)B\} = \sum_{i=1}^{n} |f(\alpha_i)\beta_i|^2,$$

for every $f(A)B$.

Since A is normal, there is a unitary matrix Φ such that

$$\bar{\Phi}'A\Phi = D = \operatorname{diag}[\alpha_1, \alpha_2, \ldots, \alpha_n],$$

where equal α_i are arranged consecutively. Thus

$$D = \alpha_1' I_1 \dot{+} \alpha_2' I_2 \dot{+} \ldots \dot{+} \alpha_k' I_k,$$

where $\alpha_i' \neq \alpha_j'$ when $i \neq j$.

Suppose that there are n_i latent roots of A equal to α_i and denote these now by α_{ij} $(j = 1, \ldots, n_i)$. Denote the corresponding roots of B in the enunciation of the theorem by β_{ij}. Let $\bar{\Phi}'B\Phi = C$, and let C be partitioned conformably with D.

We may construct the polynomials $f_i(x)$ $(i = 1, \ldots, k)$ for which

$$f_i(\alpha_i') = 1, \qquad f_i(\alpha_j') = 0, \quad \text{when } j \neq i.$$

Then $f_i(D)C = \bar{\Phi}'f_i(A)B\Phi$, and so, by hypothesis and by the definition of $f_i(x)$, we have

$$\Lambda\{f_i(D)C\} = \Lambda\{f_i(A)B\} = \sum_{h=1}^{k} |f_i(\alpha_h')|^2 \Big(\sum_{j=1}^{n_h} |\beta_{hj}|^2 \Big) = \sum_{j=1}^{n_i} |\beta_{ij}|^2. \qquad (1)$$

Also $$f_i(D) = O_1 \dot{+} \ldots \dot{+} O_{i-1} \dot{+} I_i \dot{+} O_{i+1} \dot{+} \ldots \dot{+} O_k,$$

where the O_i are null-matrices, and the I_i are unit-matrices, of order n_i. Hence, if $G_i \equiv f_i(D)C$ is partitioned conformably with D, then $G_{i,kj} = 0$, when $k \neq i$, while $G_{i,ii} = C_{ii}$. It follows that the latent roots of $f_i(D)C$ are those of C_{ii} together with zeros. By (1),

$$\Lambda(C_{ii}) = \Lambda\{f_i(D)C\} = \sum_{j=1}^{n_i} |\beta_{ij}|^2.$$

But the latent roots of $C = \bar{\Phi}'B\Phi$ are β_{ij} $(j = 1, \ldots, n_i;\ i = 1, \ldots, k)$. The lemma applies and we deduce that

$$C = C_{11} \dot{+} C_{22} \dot{+} \ldots \dot{+} C_{kk}.$$

This is sufficient for D and C to commute. Transforming by Φ, we see that A and B commute.

Suppose that α_{ij}, β_{ij} and α_{ij}, β'_{ij} are orderings of the latent roots of A and B satisfying Frobenius's theorem. The latent roots of $f_i(A)B$ are β_{ij} or β'_{ij} $(j = 1,..., n_i)$, together with $n-n_i$ zeros. Hence in Frobenius's theorem the latent roots of B associated with a set of equal latent roots of A form a unique set. This is not true in the case of our theorem. For suppose that α_{ij}, β_{ij} and α_{ij}, β'_{ij} satisfy

$$\Lambda\{f(A)B\} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}|f(\alpha_{ij})\beta_{ij}|^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}|f(\alpha_{ij})\beta'_{ij}|^2,$$

for all polynomials $f(A)B$. Then we must have

$$\sum_{j=1}^{n_i}|\beta_{ij}|^2 = \sum_{j=1}^{n_i}|\beta'_{ij}|^2 \quad (i = 1,..., k), \tag{2}$$

by putting $f(A)B = f_i(A)B$ $(i = 1,..., k)$. Conversely, if α_{ij}, β_{ij} is an ordering satisfying Theorem 1, and (2) is satisfied for some ordering α_{ij}, β'_{ij} of the latent roots of B, then

$$\begin{aligned}\Lambda\{f(A)B\} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}|f(\alpha_{ij})\beta_{ij}|^2 &= \sum_{i=1}^{k}|f(\alpha'_i)|^2\Big(\sum_{j=1}^{n_i}|\beta_{ij}|^2\Big)\\ &= \sum_{i=1}^{k}|f(\alpha'_i)|^2\Big(\sum_{j=1}^{n_i}|\beta'_{ij}|^2\Big) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}|f(\alpha_{ij})\beta'_{ij}|^2,\end{aligned}$$

and therefore the ordering α_{ij}, β'_{ij} also satisfies the theorem. I state these results in a corollary.

COROLLARY. *Let the latent roots of A be α_{ij} $(j = 1,..., n_i;\ i = 1,..., k)$, where $\alpha_{ij} = \alpha'_i$ and $\alpha'_i \neq \alpha'_j$ when $i \neq j$; and let the latent roots of B be β_{ij} or β'_{ij} $(j = 1,..., n_i;\ i = 1,..., k)$. Let A and B commute. Then, if α_{ij}, β_{ij} is an ordering satisfying Theorem* 1, *α_{ij}, β'_{ij} is another if and only if*

$$\sum_{j=1}^{n_i}|\beta_{ij}|^2 = \sum_{j=1}^{n_i}|\beta'_{ij}|^2 \quad (i = 1,..., k).$$

It is easily seen from an example that there may be various groupings of the β_{ij} into k sets of n_i members each, the sum of squared moduli of the ith set being $\sum_{j=1}^{n_i}|\beta_{ij}|^2$.

3. I shall next prove a theorem concerning a unitary matrix B. It has no analogue for general normal B.

THEOREM 2. *Let $B = V$ be unitary. Then AV is normal if and only if the sum of squared moduli of latent roots of AV is $\sum_{i=1}^{n}|\alpha_i|^2$.*

Proof. It is well known that multiplication by a unitary matrix does not alter the sum of squared moduli of elements of a matrix.

Let $C = AV$, and let the latent roots of C be γ_i $(i = 1,..., n)$. Then

$$S = \sum_{i,j=1}^{n} |c_{ij}|^2 = \sum_{i,j=1}^{n} |a_{ij}|^2 = \sum_{i=1}^{n} |\alpha_i|^2,$$

since A is normal. By Schur's theorem C is normal if and only if

$$\sum_{i=1}^{n} |\gamma_i|^2 = S,$$

and the theorem follows from $S = \sum_{i=1}^{n} |\alpha_i|^2$.

When AV is normal the polar matrices of A and those of AV commute. The latent roots of the unitary matrices V and UV are of unit modulus. Denote the latent roots of H, U, UV, by h_i, u_i, w_i $(i = 1,..., n)$ respectively. If $C = AV$, then $C = H.UV$ is a polar representation of C, and H and UV commute. By Frobenius's theorem,

$$h_i u_i = \alpha_i,$$

and

$$h_i w_i = \gamma_i;$$

whence

$$|\gamma_i| = |\alpha_i|.$$

On using Theorem 2 we obtain the non-trivial part of Corollary 1.

COROLLARY 1. *The sum of squared moduli of latent roots of AV is $\sum_{i=1}^{n} |\alpha_i|^2$ if and only if the moduli of latent roots of AV are $|\alpha_i|$ $(i = 1,..., n)$.*

Next let H be a non-negative definite Hermitian matrix, with latent roots h_i $(i = 1,..., n)$. It is known that $h_i \geqslant 0$. By Theorem 2,

$$\sum_{i=1}^{n} h_i^2 = \sum_{i=1}^{n} |\gamma_i|^2$$

if and only if $C = HV$ is normal. But C is normal if and only if $HV = VH$. Corollary 2 [cf. Parker (3)] now follows.

COROLLARY 2. *Let H be a non-negative definite Hermitian, and V be unitary. Then H and V commute if and only if the sum of squared moduli of latent roots of H is equal to the sum of squared moduli of latent roots of HV.*

It will be seen that Theorem 1 is a generalization of this corollary.

4. In Theorems 4 and 5 we shall be considering polynomials $f(A)$ whose polar representation is $f(A) = g(H)W$. The significance of this restriction is brought out by Theorem 3.

THEOREM 3. *The polar representation† of $f(A)$ is $f(A) = g(H)W$, where $g(H)$ is non-negative definite Hermitian and W is unitary, if and only if*

$$|f(\alpha_i)| = |f(\alpha_j)| = g(|\alpha_i|), \tag{3}$$

when $|\alpha_i| = |\alpha_j|$.

Proof. Let $D = \overline{\Phi}'A\Phi = \operatorname{diag}[\alpha_1,\ldots,\alpha_n]$.

Then $D = \overline{\Phi}'H\Phi.\overline{\Phi}'U\Phi$

is a polar representation of D. If

$$R = \operatorname{diag}[|\alpha_1|,\ldots,|\alpha_n|],$$

there is a diagonal matrix $Z = \operatorname{diag}[z_1,\ldots,z_n]$, $|z_i| = 1$, for which $D = RZ$ is also a polar representation of D. The Hermitian polar matrix is unique, and therefore

$$R = \overline{\Phi}'H\Phi.$$

Suppose that $f(A) = g(H)W$.

Then $f(D) = \overline{\Phi}'f(A)\Phi = \overline{\Phi}'g(H)\Phi.\overline{\Phi}'W\Phi = g(R).\overline{\Phi}'W\Phi$.

Also $$f(D) = \operatorname{diag}[f(\alpha_1),\ldots,f(\alpha_n)], \tag{4}$$

and $$g(R) = \operatorname{diag}[g(|\alpha_1|),\ldots,g(|\alpha_n|)], \tag{5}$$

while, by an argument similar to that for D and R above, it follows from

$$f(D) = g(R).\overline{\Phi}'W\Phi$$

that $$g(R) = \operatorname{diag}[|f(\alpha_1)|,\ldots,|f(\alpha_n)|]. \tag{6}$$

By comparing (5) and (6), we deduce (3).

Conversely, let us assume (3). We obtain (4) and (5), by the definition of D and R. Then (6) follows by (3). Thus $f(D) = g(R)X = Xg(R)$ is a polar representation, where $X = \operatorname{diag}[x_1,\ldots,x_n]$, $|x_i| = 1$. Hence

$$f(A) = \Phi f(D)\overline{\Phi}' = \Phi g(R)X\overline{\Phi}' = g(H)W,$$

for $$W = \Phi X\overline{\Phi}',$$

and the theorem is proved.

I shall use this theorem due to Wiegmann (**5**): '*Let A and B be normal. Then AB is normal if and only if H and B, A and V, commute.*' A version of this theorem has already been quoted in the introductory remarks.

THEOREM 4. *Let A, B and AB be normal. Then, if the Hermitian polar matrices of $f(A)$, $l(B)$ are functions of the Hermitian polar matrices of A, B respectively, $f(A)l(B)$ is normal.*

† Since the unitary matrix W is here left unrestricted we may say '*the* polar representation' meaning '*all* polar representations'.

Proof. Let $f(A) = g(H)W, \quad l(B) = m(K)X.$

By Wiegmann's theorem, H and B, A and K commute. Hence $g(H)$ and $l(B)$, $f(A)$ and $m(K)$ commute. Using Wiegmann's theorem in the reverse direction, we see that $f(A)l(B)$ is normal.

If $f(A)l(B)$ is normal, it is easily seen that the Hermitian polar matrices of $f(A)$, $l(B)$ may not be functions of H and K.

THEOREM 5. *If $f(A)l(B)$ is normal, the moduli of its latent roots are $|f(\alpha_i)l(\beta_i)|$ $(i = 1,..., n)$, for some ordering α_i, β_i.*

Proof. Let $f(A) = GW$, $l(B) = MX$ be polar representations; $f(A)$ and $l(B)$ are normal. As in Theorem 3, we may prove that the latent roots of G, M are $|f(\alpha_i)|$, $|l(\beta_i)|$ $(i = 1,..., n)$ respectively. Now $f(A)l(B)$ is normal and therefore G and M, W and M, G and X commute, provided that W and X are suitably chosen, by the version of Wiegmann's theorem quoted in the introduction. Hence†

$$f(A)l(B) = GWMX = GMWX = WXGM.$$

By Frobenius's theorem the latent roots of GM are $|f(\alpha_i)l(\beta_i)|$ for some ordering, and the latent roots of $f(A)l(B)$ are $|f(\alpha_i)l(\beta_i)|y_i$, where the y_i $(|y_i| = 1)$ are the latent roots of WX. The theorem follows.

5. The matrix AB is normal when $AB = BA$. Hence, by Theorem 1, if

$$\Lambda\{f(A)B\} = \sum_{i=1}^{n} |f(\alpha_i)\beta_i|^2$$

for every $f(A)$, then AB is certainly normal. In Theorem 6, I shall prove that it is sufficient to assume rather less for the normality of AB.

THEOREM 6. *The matrix AB is normal if and only if, for some ordering α_i, β_i $(i = 1,..., n)$, $\sum_{i=1}^{n} |f(\alpha_i)\beta_i|^2$ equals the sum of squared moduli of latent roots of every scalar polynomial $f(A)B$ for which the Hermitian polar matrix of $f(A)$ is a polynomial in the Hermitian polar matrix of A.*

Proof. Suppose that AB is normal. Then, if $f(A) = g(H)W$ and $B = KV$ are polar representations, we may use Theorem 4 with $l(B) = B$. It follows that $f(A)B$ is normal. By Theorem 5 the moduli of latent roots of $f(A)B$ are thus $|f(\alpha_i)\beta_i|$ $(i = 1,..., n)$, for some ordering α_i, β_i. Hence

$$\Lambda\{f(A)B\} = \sum_{i=1}^{n} |f(\alpha_i)\beta_i|^2. \tag{7}$$

† When $f(A)l(B)$ is normal, then $f(A)l(B) = GMWX$ for all polar matrices W and X. This is easily proved.

Suppose now that there is an ordering α_i, β_i for which (7) is satisfied for every $f(A)$ that has a polar representation $f(A) = g(H)W$. We have already proved that, if

$$\overline{\Phi}'A\Phi = D = \operatorname{diag}[\alpha_1,\ldots,\alpha_n],$$

then

$$\overline{\Phi}'H\Phi = R = \operatorname{diag}[|\alpha_1|,\ldots,|\alpha_n|].$$

On constructing the polynomial $\psi(x)$ satisfying $\psi(\alpha_i) = |\alpha_i|$ $(i = 1,\ldots,n)$, it follows that

$$\psi(D) = R,$$

and

$$\psi(A) = \Phi\psi(D)\overline{\Phi}' = \Phi R\overline{\Phi}' = H.$$

Since H is a polynomial in A, any polynomial in H is also a polynomial in A. Now let

$$f(A) = g(H)W.$$

We may suppose that $g(H) = \phi(A)$, and this implies $g(R) = \phi(D)$. From this equation and from Theorem 3 it follows that

$$\phi(\alpha_i) = g(|\alpha_i|) = |f(\alpha_i)| \quad (i = 1,\ldots,n).$$

Since, of course, $\phi(A) = g(H)$ is its own Hermitian polar matrix,

$$\Lambda\{g(H)B\} = \Lambda\{\phi(A)B\} = \sum_{i=1}^{n} |\phi(\alpha_i)\beta_i|^2 = \sum_{i=1}^{n} |g(|\alpha_i|)\beta_i|^2, \tag{8}$$

by hypothesis. We have proved (8) only for any non-negative Hermitian $g(H)$. From an inspection of the proof of Theorem 1 it is immediately clear that H and B commute if (8) is assumed only for the polynomials $g_i(H)$ defined for H in the same way as $f_i(A)$ is defined for A. These polynomials $f_i(A)$ used in the proof of Theorem 1 are non-negative definite Hermitian, being normal matrices with non-negative latent roots. Hence the corresponding $g_i(H)$ are also non-negative definite Hermitian; (8) holds for them and the commutativity of H and B follows.

If $\psi(A) = H$, we have the polar representations $\psi(A) = HI$ and $A = HU$. Hence, by the hypothesis,

$$\Lambda(HB) = \sum_{i=1}^{n} |\alpha_i\beta_i|^2$$

and

$$\Lambda(AB) = \sum_{i=1}^{n} |\alpha_i\beta_i|^2.$$

We deduce that $\Lambda(AB) = \Lambda(HB)$, where HB is normal, since H and B commute. On using Theorem 2 we see that AB is normal. The theorem is thus proved.

A corollary may be obtained by an argument similar to that which led to the corollary of Theorem 1.

COROLLARY. *Let the latent roots of A be α_{ij} $(j = 1,..., m_i;\ i = 1,..., r)$, where $|\alpha_{ij}| = |\alpha_i''|$ and $|\alpha_i''| \neq |\alpha_j''|$ when $i \neq j$; and let the latent roots of B be β_{ij} or β'_{ij} $(j = 1,..., m_i;\ i = 1,..., r)$. Let AB be normal. Then, if α_{ij}, β_{ij} is an ordering satisfying Theorem 6, α_{ij}, β'_{ij} is another if and only if*

$$\sum_{j=1}^{m_i} |\beta_{ij}|^2 = \sum_{j=1}^{m_i} |\beta'_{ij}|^2 \quad (i = 1,..., r).$$

6. We may remark in conclusion, without proof, that the polynomials $f_i(A)$ and $g_i(H)$ are the principal idempotent elements of A and H respectively [cf. Drazin (**1**)].

In this connexion, as some work in the proof of Theorem 6 indicates, the condition

$$\Lambda\{f_i(A)B\} = \sum_{j=1}^{n_i} |\beta_{ij}|^2 \quad (i = 1,..., k)$$

is a sufficient guarantee of the commutativity of A and B. Analogously, the normality of AB is ensured by the condition

$$\Lambda\{g_i(H)UB\} = \sum_{j=1}^{m_i} |\beta_{ij}|^2 \quad (i = 1,..., r),$$

for *some* polar representation $A = HU$, though this requires much fuller amplification.

REFERENCES

1. M. P. Drazin, 'On diagonable and normal matrices', *Quart. J. of Math.* (Oxford) (2), 2 (1951), 189–98.
2. M. P. Drazin, J. W. Dungey, and K. W. Gruenberg, 'Some theorems on commutative matrices', *J. of London Math. Soc.* 26 (1951), 221–8.
3. W. V. Parker, 'The characteristic roots of matrices', *Duke Math. J.* 12 (1945), 519–26.
4. I. Schur, 'Über die charakterischen Wurzeln einer linearen Substitution mit einer Anwendung auf die Theorie der Integralgleichungen', *Math. Annalen*, 66 (1909), 488–510.
5. N. A. Wiegmann, 'Normal products of matrices', *Duke Math. J.* 15 (1948), 633–8.
6. J. Williamson, 'Note on a principal-axis transformation of non-Hermitian matrices', *Bull. American Math. Soc.* 45 (1939), 920–2.
7. A. Wintner and F. D. Murnaghan, 'On a polar representation of non-singular matrices', *Proc. National Academy Sciences, U.S.A.* 17 (1931), 676–8.

THE REDUCTION BY ONE OF THE NUMBER OF INDEPENDENT VARIABLES IN SOME SYSTEMS OF PARTIAL DIFFERENTIAL EQUATIONS

By A. J. A. MORGAN (*Pasadena*)

[Received 23 October 1951]

1. Introduction

BIRKHOFF has suggested a method whereby the number of independent variables of some partial differential equations can be reduced by one. His proposed method is illustrated by several examples (**1**), (**2**).

The work presented herein constitutes a generalization of Birkhoff's suggestion, with some essential differences, to systems of arbitrary partial differential equations and provides a rigorous proof of its validity.

Let $\Phi_\delta = 0$ ($\delta = 1,\ldots, n$) be a system of arbitrary partial differential equations. Let G_1 be an arbitrary continuous one-parameter group of transformations, of a form to be specified later, of the dependent and independent variables of the equations $\Phi_\delta = 0$. It will be shown that, if each of the equations in the system $\Phi_\delta = 0$ is invariant† under G_1, then the number of its independent variables can be reduced by one.

Two examples, one from classical elasticity and one from fluid mechanics, are given to illustrate the theory.

2. The one-parameter continuous group of transformations to be considered

The succeeding investigation will be concerned with a one-parameter continuous group of transformations of the form:

$$G_1\colon \begin{cases} \bar{x}^i = f^i(x^1,\ldots, x^m; a) & (i = 1,\ldots, m;\ m \geqslant 2), \\ \bar{y}_\delta = f_\delta(y_\delta; a) & (\delta = 1,\ldots, n;\ n \geqslant 1), \end{cases} \tag{1}$$

where a is a numerical parameter and the transformations $x^i \to \bar{x}^i$ form a sub-group S_{G_1} of G_1. Subsequently the x^i and y_δ will, respectively, be identified with the independent and dependent variables of a system of partial differential equations.

Consider now the y_δ to be differentiable functions of the x^i up to any required order. If the transformations of the partial derivatives of the y_δ with respect to the x^i are appended to those of (1), it can be shown that the resulting set of transformations also forms a continuous one-parameter group. The new groups constructed in this manner will be

† The meaning of this terminology will be made precise in Definition 4.

called *enlargements* of the group G_1 and will be denoted by $G_1^{E_1},\ldots, G_1^{E}$ according as the transformations among the first, second,..., kth partial derivatives of the y_δ are added successively to those of $G_1,\ldots, G_1^{E_{k-1}}$.

By elementary group theory (3) the group† (1) has $m+n-1$ functionally independent (when considered as functions of the $m+n$ independent variables $x^1,\ldots, x^m, y_1,\ldots, y_n$) absolute invariants

$$\eta_1(x^1,\ldots, x^m), \quad \ldots, \quad \eta_{m-1}(x^1,\ldots, x^m),$$
$$g_1(y_1,\ldots, y_n, x^1,\ldots, x^m), \quad \ldots, \quad g_n(y_1,\ldots, y_n, x^1,\ldots, x^m), \tag{2}$$

and it is possible to choose them so that the Jacobian

$$\frac{\partial(g_1,\ldots, g_n)}{\partial(y_1,\ldots, y_n)} \neq 0, \tag{3}$$

and

$$R\left(\frac{\partial(\eta_1,\ldots, \eta_{m-1})}{\partial(x^1,\ldots, x^m)}\right) = m-1, \tag{4}$$

where R denotes the rank of the indicated Jacobian matrix.

Suppose that we now consider the y_δ and $\bar{y}_\delta$ to be implicitly defined as functions of the x^i and $\bar{x}^i$ (respectively) by the equations‡

$$z_\delta(x^1,\ldots, x^m) = g_\delta(y_1,\ldots, y_n, x^1,\ldots, x^m) \tag{5}$$

and

$$\bar{z}_\delta(\bar{x}^1,\ldots, \bar{x}^m) = g_\delta(\bar{y}_1,\ldots, \bar{y}_n, \bar{x}^1,\ldots, \bar{x}^m), \tag{6}$$

where the g_δ are the absolute invariants of the group (1) indicated in (2). The following result can be shown to hold.

THEOREM 1. *A necessary and sufficient condition for the y_δ, implicitly defined as functions of $x^1,\ldots, x^m$ by the relations* (5), *to be exactly the same functions of $x^1,\ldots, x^m$ as the $\bar{y}_\delta$, implicitly defined as functions of $\bar{x}^1,\ldots, \bar{x}^m$ by the relations* (6), *are of the $\bar{x}^1,\ldots, \bar{x}^m$ is that*

$$z_\delta(x^1,\ldots, x^m) = \bar{z}_\delta(\bar{x}^1,\ldots, \bar{x}^m) = z_\delta(\bar{x}^1,\ldots, \bar{x}^m). \tag{7}$$

The condition (7) *can be replaced by*

$$z_\delta(x^1,\ldots, x^m) = F_\delta(\eta_1,\ldots, \eta_{m-1}), \tag{8}$$

where the η's are the set of absolute invariants of the sub-group $x^i \to \bar{x}^i$ of the group (1).

Proof. The necessity of the condition will first be proved. The g_δ are absolute invariants of the group (1); hence, under its transformations, they must satisfy the relations

$$g_\delta(y_1,\ldots, y_n, x^1,\ldots, x^m) = g_\delta(\bar{y}_1,\ldots, \bar{y}_n, \bar{x}^1,\ldots, \bar{x}^m) \tag{9}$$

when the $y_1,\ldots, y_n, x^1,\ldots, x^m$ are considered as independent variables.

† A continuous one-parameter group of transformations will sometimes be referred to as, simply 'a group'.

‡ Unless otherwise indicated, Greek and Latin indices will have the ranges 1,..., n and 1,..., m respectively.

Now consider the y_δ and $\bar{y}_\delta$ to be functions of the x^i and $\bar{x}^i$ respectively. Since, by hypothesis, the y_δ are exactly the same functions of the x^i that the $\bar{y}_\delta$ are of the $\bar{x}^i$, the g_δ, when considered as functions of the x^i, must also have the same property. Hence on combining the relations (5), (6), (9) it follows that

$$z_\delta(x^1,\ldots, x^m) = \bar{z}_\delta(\bar{x}^1,\ldots, \bar{x}^m) = z_\delta(\bar{x}^1,\ldots, \bar{x}^m)$$

under the transformations of the group (1).

To prove the sufficiency, proceed as follows. As a consequence of (3) the relations (5) and (6) can be inverted for the y_δ and $\bar{y}_\delta$ to give, respectively,

$$y_\delta(x^1,\ldots, x^m) = h_\delta(z_1,\ldots, z_n, x^1,\ldots, x^m) \tag{10}$$

and

$$\bar{y}_\delta(\bar{x}^1,\ldots, \bar{x}^m) = \bar{h}_\delta(\bar{z}_1,\ldots, \bar{z}_n, \bar{x}^1,\ldots, \bar{x}^m) \tag{11}$$

in some neighbourhood of the $(x^1,\ldots,x^m)$ and $(\bar{x}^1,\ldots,\bar{x}^m)$ spaces respectively.

The g_δ are, under the transformations of the group (1), exactly the same functions of the y_δ and x^i as they are of the $\bar{y}_\delta$ and $\bar{x}^i$. Hence, the h_δ must be exactly the same functions of the z_δ and x^i as the $\bar{h}_\delta$ are of the $\bar{z}_\delta$ and $\bar{x}^i$. But, by hypothesis, the z_δ are exactly the same functions of the x^i as the $\bar{z}_\delta$ are of the $\bar{x}^i$. Therefore, by (10), (11), and the previous arguments, the y_δ must be exactly the same functions of the x^i as the $\bar{y}_\delta$ are of the $\bar{x}^i$.

The relations (7) imply that the z_δ are absolute invariants of the subgroup S_{G_1}. By elementary group theory every absolute invariant of S_{G_1} must be a function of the $m-1$ functionally independent absolute invariants $\eta_1,\ldots, \eta_{m-1}$; hence, the condition (7) can be replaced by (8), and this completes the proof of the theorem.

In view of the above result, on identifying the y_δ and x^i with the dependent and independent variables of a system of partial differential equations respectively, it is convenient to define:

DEFINITION 1. By *invariant solutions of a system of partial differential equations* is meant that class of solutions of a system of partial differential equations which have the property that the y_δ are exactly the same functions of the x^i as the $\bar{y}_\delta$ are of the $\bar{x}^i$.

When considering invariant solutions of systems of partial differential equations, by Theorem 1 it is only necessary to consider the relations (5) or (6) to be given in the form

$$F_\delta(\eta_1,\ldots, \eta_{m-1}) = g_\delta(y_1,\ldots, y_n, x^1,\ldots, x^m), \tag{12}$$

where the functions $F_\delta(\eta_1,\ldots, \eta_{m-1})$ are at least in class $C^{(1)}$.

By (3) and the assumed continuity and differentiability properties of

the functions F_δ and g_δ it can be shown that the conditions of the implicit function theorem are satisfied by (12). Hence, the y_δ can be expressed as functions of the F_δ and x^i in some neighbourhood of the $(x^1,\ldots, x^m)$ space, say,

$$y_\delta = \lambda_\delta(F_1,\ldots, F_n, x^1,\ldots, x^m). \tag{13}$$

3. Conformally invariant differential forms

DEFINITION 2. By a *differential form of the k-th order in m independent variables* is meant a function, usually in class $C^{(1)}$ or greater, of the form

$$\Phi\left(x^1,\ldots, x^m, y_1,\ldots, y_n,\ldots, \frac{\partial^k y_1}{\partial (x^1)^k},\ldots, \frac{\partial^k y_n}{\partial (x^m)^k}\right) \tag{14}$$

whose arguments are the variables $x^1,\ldots, x^m$, functions $y_1,\ldots, y_n$ dependent on them, and the partial derivatives of the y_δ with respect to the x^i up to the kth order.

Suppose that each of the arguments in (14) transforms under the transformation laws of a continuous one-parameter group with symbol V [(3) 28]† and numerical parameter a. Each of the arguments in (14) can then be considered as an independent variable of the group with symbol V. If there are p such arguments, there will be no loss in generality in calling them $z^1,\ldots, z^p$.

DEFINITION 3. A differential form Φ will be said to be *conformally invariant under a one-parameter group* G_1, if, under the transformations of the group, it satisfies the relation

$$\Phi(\bar{z}^1,\ldots, \bar{z}^p) = F(z^1,\ldots, z^p; a)\Phi(z^1,\ldots, z^p), \tag{15}$$

where Φ is exactly the same function of the z's as it is of the $\bar{z}$'s and F is some function of the z's and the parameter a.

It will be said that Φ is *constant conformally* invariant if under the transformations of the group with symbol V the relation (15) is satisfied with the F a function of a only; Φ will, under the same conditions, be said to be *absolutely invariant* if the relation (15) is satisfied with F identically equal to one.

THEOREM 2. *If* Φ *of* (14) *is at least in class* $C^{(1)}$ *with respect to each of its arguments, then a necessary and sufficient condition for* Φ *to be conformally invariant under a continuous one-parameter group of transformations with symbol* V *is that*

$$V\Phi = \omega(z^1,\ldots, z^p)\Phi(z^1,\ldots, z^p) \tag{16}$$

for some $\omega(z^1,\ldots, z^p)$.

† The *symbol* of a one-parameter group is a first-order linear partial differential operator, and in this case

$$V = \xi^1(z^1,\ldots, z^p)\frac{\partial}{\partial z^1} + \cdots + \xi^p(z^1,\ldots, z^p)\frac{\partial}{\partial z^p}.$$

The proof of the above statement is similar to that given by Eisenhart (3) for the case of absolute invariance and need not be repeated here.

The condition (16) of Theorem 2 can be replaced by

$$\Phi(z^1,\ldots, z^p) = e^{\zeta(z^1,\ldots,z^p)}\Phi_0(z^1,\ldots, z^p), \tag{17}$$

where Φ_0 is a general absolute invariant of the group with symbol V and $\zeta(z^1,\ldots, z^p)$ is a determinable function of $z^1,\ldots, z^p$. The result (17) can be obtained by solving the first-order partial differential equation (16) by standard methods (4) and supposing that the solution of the equation $V\Phi = 0$ is known. The condition (17), rather than (16), will be most useful in the succeeding investigation.

DEFINITION 4. It will be said that a *system of k-th order partial differential equations* $\Phi_\delta = 0$ *is invariant under a continuous one-parameter group of transformations* G_1 if each of the kth order differential forms Φ_δ is conformally invariant under the transformations of $G_1^{E_k}$.

4. Systems of partial differential equations invariant under continuous one-parameter groups of transformations

Consider a system of partial differential equations of the kth order in m ($\geqslant 2$) independent variables $x^1,\ldots, x^m$ and n ($\geqslant 1$) dependent variables $y_1,\ldots, y_n$ of the form:

$$\Phi_\delta\left(x^1,\ldots, x^m, y_1,\ldots, y_n,\ldots, \frac{\partial^k y_1}{\partial(x^1)^k},\ldots, \frac{\partial^k y_n}{\partial(x^m)^k}\right) = 0, \tag{18}$$

where each of the differential forms is conformally invariant under the kth enlargement of the group G_1.

The question to be answered is: Can the *invariant solutions* of the system (18) be expressed in terms of the solutions of a system of the kth order partial differential equations in $m-1$ independent variables? The following result shows that the answer is in the affirmative.

THEOREM 3. *If each of the differential forms* $\Phi_1,\ldots, \Phi_n$ *in the system of partial differential equations* (18) *is conformally invariant under the k-th enlargement of the group* (1), *then the invariant solutions of* (18) *can be expressed in terms of the solutions of a system of the form*

$$A_\delta\left(\eta_1,\ldots, \eta_{m-1}, F_1,\ldots, F_n,\ldots, \frac{\partial^k F_1}{\partial\eta_1^k},\ldots, \frac{\partial^k F_n}{\partial\eta_{m-1}^k}\right) = 0, \tag{19}$$

a system of k-th order partial differential equations in $m-1$ *independent variables* $\eta_1,\ldots, \eta_{m-1}$. *In the above the* η*'s are the* $m-1$ *absolute invariants of the group* S_{G_1} *and the functions* F_δ, *at least in class* $C^{(k)}$, *are those of relations* (12).

Proof. By (4) it is possible to express $m-1$ of the x's in terms of $\eta_1,\ldots,\eta_{m-1}$ and the remaining x, say ${}^*x^m$. Hence, it is possible to write

$$ {}^*x^i = h^i(\eta_1,\ldots,\eta_{m-1}, {}^*x^m) \quad (i = 1,\ldots,m-1). \tag{20} $$

By (20) the relations (13) can be expressed as:

$$ y_\delta = \Omega_\delta(F_1,\ldots,F_n,\eta_1,\ldots,\eta_{m-1}, {}^*x^m). \tag{21} $$

The kth order partial derivatives of the y_δ with respect to the x's can now be written as

$$ y_{\alpha,j_1\ldots j_k} = \Omega_{\alpha j_1\ldots j_k}\left(\frac{\partial^k F_\delta}{\partial\eta_{i_1}\ldots\partial\eta_{i_k}},\ldots,\frac{\partial F_\delta}{\partial\eta_{i_1}}, F_\delta, \eta_{i_1}, {}^*x^m\right), \tag{22} $$

where

$$ \alpha,\delta = 1,\ldots,n, \quad i_1,\ldots,i_k = 1,\ldots,m-1, \quad j_1,\ldots,j_k = 1,\ldots,m $$

and the notation indicates that the functions $\Omega_{\alpha j_1\ldots j_k}$ may depend on all of the variables $\dfrac{\partial^k F_\delta}{\partial\eta_{i_1}\ldots\partial\eta_{i_k}},\ldots,\dfrac{\partial F_\delta}{\partial\eta_{i_1}}$,† F_δ, η_{i_1}, and ${}^*x^m$.

The substitution of (20), (21), and (22) into each of the differential forms Φ_δ of (18) yields the relations

$$ \Phi_\delta = B_\delta\left({}^*x^m,\eta_1,\ldots,\eta_{m-1},F_1,\ldots,F_n,\ldots,\frac{\partial^k F_1}{\partial\eta_1^k},\ldots,\frac{\partial^k F_n}{\partial\eta_{m-1}^k}\right). \tag{23} $$

By using the condition (17) of Theorem 2 each of the differential forms Φ_δ can always be expressed as

$$ \Phi_\delta = e^{\zeta_\delta}{}_\delta\Phi_0, \tag{24} $$

where the functions ${}_\delta\Phi_0$ are absolute invariants of the group $G_1^{E_k}$ and the arguments of the functions of ζ_δ and ${}_\delta\Phi_0$ are exactly the same as those of Φ_δ.

Similarly, by (23) and on using (20), (21), (22) in (24), the following expression is obtained

$$ B_\delta = e^{\sigma_\delta}{}_\delta\phi_0, \tag{25} $$

where the functions ${}_\delta\phi_0$, when considered in terms of the variables of $G_1^{E_k}$, are absolute invariants of the group $G_1^{E_k}$ and, in particular,

$$ {}_\delta\Phi_0\left(x^1,\ldots,x^m,y_1,\ldots,y_n,\ldots,\frac{\partial^k y_1}{\partial(x^1)^k},\ldots,\frac{\partial^k y_n}{\partial(x^m)^k}\right) $$
$$ = {}_\delta\phi_0\left({}^*x^m,\eta_1,\ldots,\eta_{m-1},F_1,\ldots,F_n,\ldots,\frac{\partial^k F_1}{\partial\eta_1^k},\ldots,\frac{\partial^k F_n}{\partial\eta_{m-1}^k}\right) \tag{26} $$

under the substitutions (20), (21), (22). The arguments of the function σ_δ are exactly the same as those of B_δ.

† Each of these quantities, when considered in terms of the variables of $G_1^{E_k},\ldots,G_1^{E_1}$ respectively, can be shown to be absolute invariants of the groups $G_1^{E_k},\ldots,G_1^{E_1}$.

It remains to be shown that each of the differential forms ${}_\delta\phi_0$ is independent of ${}^*x^m$. The following argument is used. Let the y_δ be any set of functions of the x^i having the invariance property defined in Theorem 1. Then

$$ {}_\delta\Phi_0 = {}_\delta\psi_0(x^1,\ldots, x^m), \tag{27}$$

and, under the transformations of $G_1^{E_k}$,

$$ {}_\delta\Phi_0\left(\bar{x}^1,\ldots, \bar{x}^m, \bar{y}_1,\ldots, \bar{y}_n,\ldots, \frac{\partial^k \bar{y}_1}{\partial(\bar{x}^1)^k},\ldots, \frac{\partial^k \bar{y}_n}{\partial(\bar{x}^m)^k}\right) = {}_\delta\psi_0(\bar{x}^1,\ldots, \bar{x}^m), \tag{28}$$

where the same functional notation ${}_\delta\psi_0$ has been used in (27) and (28) since the y_δ are exactly the same functions of the x^i as the $\bar{y}_\delta$ are of the $\bar{x}^i$. The ${}_\delta\Phi_0$ are absolute invariants of the group $G_1^{E_k}$. Hence

$$ {}_\delta\psi_0(x^1,\ldots, x^m) = {}_\delta\psi_0(\bar{x}^1,\ldots, \bar{x}^m). \tag{29}$$

The relations (29) imply that the ${}_\delta\psi_0$ are absolute invariants of the group S_{G_1}; therefore

$$ {}_\delta\psi_0(x^1,\ldots, x^m) = {}_\delta\Psi_0(\eta_1,\ldots, \eta_{m-1}): \tag{30}$$

that is, the ${}_\delta\psi_0$ must, by elementary group-theory, always be expressible as functions of the $m-1$ absolute invariants $\eta_1,\ldots, \eta_{m-1}$ of S_{G_1}.

Similarly, since the F_δ are (arbitrary) functions of the η's, it follows that

$$ {}_\delta\phi_0\left({}^*x^m, \eta_1,\ldots, \eta_{m-1}, F_1,\ldots, F_n,\ldots, \frac{\partial^k F_1}{\partial \eta_1^k},\ldots; \frac{\partial^k F_n}{\partial \eta_{m-1}^k}\right) = {}_\delta\phi_0'({}^*x^m, \eta_1,\ldots, \eta_{m-1}). \tag{31}$$

All of the arguments in the ${}_\delta\phi_0$ of the above equations can assume arbitrary values; hence equations (31) can be treated as identities.

The equations (26), (30), (31) imply that

$$ {}_\delta\Psi_0(\eta_1,\ldots, \eta_{m-1}) = {}_\delta\phi_0'({}^*x^m, \eta_1,\ldots, \eta_{m-1}). \tag{32}$$

But, the equations (32) imply that the ${}_\delta\phi_0'$ must be independent of ${}^*x^m$ since otherwise the ${}_\delta\psi_0$ could not be absolute invariants of the group S_{G_1}. On retracing the steps in the above argument it is seen that the ${}_\delta\phi_0$ of (25) must be independent of ${}^*x^m$. Hence, the equations (18) can be expressed as

$$ {}_\delta\phi_0 = A_\delta\left(\eta_1,\ldots, \eta_{m-1}, F_1,\ldots, F_n,\ldots, \frac{\partial^k F_1}{\partial \eta_1^k},\ldots, \frac{\partial^k F_n}{\partial \eta_{m-1}^k}\right) = 0 \tag{33}$$

since the quantities e^{σ_δ} of (25) are not identically zero for arbitrary values of the indicated arguments. This completes the proof of Theorem 3.

The solutions of the system (33) are then related to those of (18) by the equation (12). The solutions of (18) determined in such a manner are invariant, in the sense of Definition 1, under the same group which leaves the system of partial differential equations invariant.

Consider that another set of functionally independent absolute invariants of S_{G_1} be chosen; say

$$\zeta_i = f_i(\eta_1, \ldots, \eta_{m-1}), \qquad J(\zeta_i/\eta_i) \neq 0,$$

and that the F_δ of (12) are replaced by functions G_δ of the ζ_i. It is evident that the result of Theorem 3 holds; that is, the resulting system of equations will have the ζ_i and G_δ as independent and dependent variables respectively. This system of partial differential equations is related to the system (33) by the transformations of independent variable

$$\zeta_i = f_i(\eta_1, \ldots, \eta_{m-1}).$$

5. Illustration of the theory

Two examples, one from classical elasticity and one from fluid mechanics, will be used to illustrate the theory developed in the preceding sections.

Example I. The deflexions of plates of uniform thickness

In the classical theory the deflexions w of constant-thickness plates are governed by the fourth-order partial differential equation

$$\nabla^4 w = \frac{\partial^4 w}{\partial x^4} + 2\frac{\partial^4 w}{\partial x^2 \partial y^2} + \frac{\partial^4 w}{\partial y^4} = \frac{P(x,y)}{D}, \tag{34}$$

where $P(x, y)$ is the loading (force/unit area) on the plate, D is a constant known as the *plate-stiffness factor*, and x, y are coordinates in the middle (undeflected) surface of the plate.

The equation (34) is absolutely invariant under the continuous one-parameter group

$$\bar{x} = a^m x, \qquad \bar{y} = a^n y, \qquad \bar{w} = a^p w, \tag{35}$$

where a is the numerical parameter ($\neq 0$), m, n, p are real numbers, if

$$\frac{n}{m} = 1, \qquad \frac{p}{m} = 4, \tag{36}$$

and

$$P(\bar{x}, \bar{y}) = P(x, y). \tag{37}$$

Two absolute invariants of the group (35), corresponding to the functions η and g of (2), are

$$yx^{-n/m} \quad \text{and} \quad wx^{-p/m}.$$

Hence, by (12) and (36), the invariant solutions of (34) must be of the form

$$w = x^4 f(\eta) \quad \text{with } \eta = y/x. \tag{38}$$

By Theorem 3, $f(\eta)$ must satisfy an ordinary differential equation, obtained by the substitution of (38) in (34): that is

$$(\eta^2+1)^2\frac{d^4f}{d\eta^4}-4\eta(\eta^2+1)\frac{d^3f}{d\eta^3}+4(3\eta^2+1)\frac{d^2f}{d\eta^2}-24\eta\frac{df}{d\eta}+24f=\frac{P(\eta)}{D}, \tag{39}$$

where the condition (37) implies that P must be a function of η. Equation (39) can be used to determine the deflexions of a uniformly loaded ($P(\eta)$ = constant) semi-infinite wedge-shaped plate. The results of such an investigation are reported elsewhere (**5**).

Example II: The laminar boundary layer

As a second example consider a system of two partial differential equations which arise in the classical boundary-layer theory of Blasius (**6**),

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=0, \qquad u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}=\nu\frac{\partial^2 u}{\partial y^2}. \tag{40}$$

The equations† (40) are constant conformally invariant under the group

$$\bar{x}=a^m x, \qquad \bar{y}=a^n y, \qquad \bar{u}=a^p u, \qquad \bar{v}=a^q v, \tag{41}$$

where a is the numerical parameter ($\neq 0$), m, n, p, q are real numbers, if

$$n/m=\tfrac{1}{2}, \qquad q/m=\tfrac{1}{2} \tag{42}$$

and

$$p=0. \tag{43}$$

The condition (43), equivalent to $u=\bar{u}$, is imposed from considerations of the physical nature of the laminar boundary-layer.

One set of absolute invariants of the group (41), corresponding to the functions η, g_1, g_2 of the general theory, is

$$yx^{-n/m}, \quad ux^{-p/m}, \quad vx^{-q/m}. \tag{44}$$

Hence, by (12), (42), and (43), the substitution of

$$u=F_1(\eta), \qquad v=x^{-\frac{1}{2}}F_2(\eta)$$

with $\eta=yx^{-\frac{1}{2}}$, into (40) will (by Theorem 3) yield the system of simultaneous ordinary differential equations‡

$$-\tfrac{1}{2}\eta\frac{dF_1}{d\eta}+\frac{dF_2}{d\eta}=0, \qquad -\tfrac{1}{2}\eta F_1\frac{dF_1}{d\eta}+F_2\frac{dF_1}{d\eta}=\nu\frac{d^2F_1}{d\eta^2}. \tag{45}$$

† In practice the first of equations (40) is identically satisfied by defining a stream function ψ such that $u=\partial\psi/\partial y$, $v=-\partial\psi/\partial x$. A single partial differential equation is obtained with ψ as the dependent variable. An application of the above theory will yield an ordinary non-linear differential equation. Example II has been given in order to illustrate the theory in the case of systems of partial differential equations.

‡ P. Lagerstrom has derived these equations, in some unpublished lecture notes, by using Birkhoff's method.

6. Conclusion

The substitutions, previously given without rigorous mathematical foundation, used to obtain the invariant solutions of partial differential equations arising in applied mathematics can easily be obtained by means of the theory presented herein. Examples of this are given by the problem of the boundary layer over a flat plate with pressure gradient produced by an external velocity of the form $U = Ax^q$ (A, q real constants) (**7**) and by the problem of the spread of a cylindrical jet (**8**). In fact most of the known cases, including the two mentioned above, can be motivated by using groups of the form (41).

The method presented above does not depend on the boundary conditions which the solutions of a given partial differential equation must satisfy. Thus, for a particular problem, the invariant solutions of the equation may not satisfy the prescribed boundary conditions. Nevertheless, if from physical considerations it is suspected that an invariant solution will satisfy the boundary conditions, the theory developed herein enables one to find its form with relative ease. Once this is done it is a simple matter to check whether such a solution will (or will not) satisfy the boundary conditions. Conversely, it may be possible to obtain new physically significant solutions of the partial differential equations arising in physical theories by examining them for invariant solutions without initial regard to boundary conditions.

The author wishes to express his deep appreciation for the valuable suggestions contributed by Professor A. D. Michal of the California Institute of Technology.

REFERENCES

1. G. Birkhoff, 'Dimensional analysis of partial differential equations', *Electrical Engineering*, 67 (1948), 1185–8.
2. —— *Hydrodynamics* (Princeton, 1950), Chapter IV.
3. L. P. Eisenhart, *Continuous Groups of Transformations* (Princeton, 1933).
4. A. Cohen, *Differential Equations* (New York, 2nd ed., 1933), 252.
5. A. J. A. Morgan, 'Uniformly loaded semi-infinite wedge-shaped plates', *J. of Aeronaut. Sci.* 18 (1951), 845.
6. H. Blasius, 'Grenzschichten in Flüssigkeiten mit kleiner Reibung', *Zeits. für Math. u. Phys.* 56 (1908), 1–37.
7. V. M. Falkner and S. W. Skan, 'Some approximate solutions of the boundary-layer equations', *Aeronautical Research Committee* [Great Britain] *Reports and Memoranda*, No. 1314 (1930).
8. H. Schlichting, 'Laminare Strahlausbreitung', *Zeits. für angew. Math. u. Mech.* 13 (1933), 260–3.

FUNCTION SPACE: HILBERT-SPACE CORRESPONDENCES IN QUANTUM MECHANICS

By C. R. PUTNAM (*Purdue University, Indiana*)

[Received 12 December 1951]

LET L^2 denote the space of quadratically integrable functions f on the line $-\infty < x < \infty$ and let $\mathfrak{H}$ denote the Hilbert space of vectors X with components $x_1, x_2, \ldots$, subject to $\sum |x_k|^2 < \infty$. It is well known that, in terms of any orthonormal, complete sequence of functions $\phi_1, \phi_2, \ldots$ in L^2, the Fischer–Riesz theorem furnishes a one-to-one correspondence between L^2 and $\mathfrak{H}$. Let X denote the vector in $\mathfrak{H}$ corresponding to the function f in L^2 and let $Q = (Q_{jk})$ denote the 'coordinate' matrix with respect to the system $\{\phi_k\}$. It is easily shown that QX is in $\mathfrak{H}$ and corresponds to xf whenever the latter function belongs to L^2. On the other hand, it is quite possible, as will be shown by an example, that QX can belong to $\mathfrak{H}$ even if xf does not belong to L^2. For the case in which the ϕ_k are the *eigenfunctions* of the quantum-mechanical, linear-oscillator, boundary-value problem, however, it will be shown that this latter situation cannot occur; thus, in this case, QX cannot be in $\mathfrak{H}$ *unless* xf is in L^2. Certain consequences of the above results, dealing with the moments of distribution functions, will be obtained. Similar results are valid if Q is replaced by the momentum matrix P.

Let $\phi_1, \phi_2, \ldots$ denote an orthonormal, complete sequence of functions on $-\infty < x < \infty$ for which all functions $x\phi_k$ belong to L^2, so that

$$\int_{-\infty}^{\infty} |x\phi_k|^2 \, dx < \infty, \tag{1}$$

and define the coordinate matrix Q by

$$Q = (Q_{jk}) = \left(\int_{-\infty}^{\infty} x\phi_k \phi_j^* \, dx \right). \tag{2}$$

Let f denote an arbitrary function of L^2, so that

$$\int_{-\infty}^{\infty} |f|^2 \, dx < \infty, \tag{3}$$

Quart. J. Math. Oxford (2), 3 (1952), 260–7.

and define its Fourier coefficients x_k, which constitute the components of the vector X, by

$$x_k = \int_{-\infty}^{\infty} f\phi_k^* \, dx. \tag{4}$$

The Parseval relation

$$\sum |x_k|^2 = \int_{-\infty}^{\infty} |f|^2 \, dx \quad (< \infty) \tag{5}$$

implies that X is in $\mathfrak{H}$. Furthermore, an application of the polarized form of the Parseval relation yields, in virtue of (2) and (4),

$$(QX)_k = \sum_n Q_{kn} x_n = \int_{-\infty}^{\infty} xf\phi_k^* \, dx.$$

Hence, in case the function xf is in L^2, the vector QX must be in $\mathfrak{H}$; cf. (5).

Let $\xi_1, \xi_2, \ldots$ denote the normalized *eigenfunctions* associated with the linear oscillator† so that

$$\xi_k = c_k H_k(x) e^{-\frac{1}{2}x^2} \quad (= \xi_k^*).$$

That (1) is fulfilled is clear; moreover, the corresponding coordinate matrix Q defined by (2), for $\phi_k = \xi_k$, becomes

$$Q = (Q_{jk}) = 2^{-\frac{1}{2}}(j^{\frac{1}{2}}\delta_{j+1,k} + k^{\frac{1}{2}}\delta_{j,k+1}) \tag{6}$$

(at least to within a constant multiple which can be supposed to be 1). The following will be shown:

(I) *Let f be an arbitrary function of L^2 and define its Fourier coefficients x_k, in terms of $\phi_k = \xi_k$, by* (4). (*Then the vector $X = \{x_k\}$ is in $\mathfrak{H}$.*) *Suppose also that Q is defined by* (6) *and that the vector QX belongs to $\mathfrak{H}$. Then necessarily the function xf belongs to L^2; thus xf satisfies*

$$\int_{-\infty}^{\infty} |xf|^2 \, dx < \infty \tag{7}$$

(*and hence QX corresponds to xf*).

In order to prove (I) it will be convenient to deduce first the following lemma:

(A) *Let f and g be in L^2 and suppose that the vectors $X = \{x_k\}$ and $Y = \{y_k\}$ and the matrix Q are defined by* (4),

$$y_k = \int_{-\infty}^{\infty} g\xi_k^* \, dx,$$

and (2), *for $\phi_k = \xi_k$.*

† E. C. Kemble, *The Fundamental Principles of Quantum Mechanics* (New York and London, 1937), 370–1.

Suppose that QX and QY (in addition to X and Y) belong to $\mathfrak{H}$. Then each of the iterated series

$$\sum_{k=1}^{\infty}\Big(\sum_{j=1}^{\infty} y_j^* q_{jk}\Big)x_k \quad \text{and} \quad \sum_{k=1}^{\infty}\Big(\sum_{j=1}^{\infty} q_{kj}x_j\Big)y_k^*$$

is convergent to the same (finite) sum.

If (A, B) denotes the scalar product $\sum_k a_k b_k^*$ of two vectors $A = \{a_k\}$ and $B = \{b_k\}$, the assertion of (A) is that $(X, QY) = (QX, Y)$. It should be noted that Q is not a bounded matrix† and that the last equation is *not* a consequence of the 'formally Hermitian' nature of Q (namely, that $q_{jk} = q_{kj}^*$). It is clear that

$$(X, QY)-(QX, Y) = \lim_{n\to\infty}\sum_{k=1}^{n}\Big\{\sum_{j=k+1}^{\infty}(q_{jk}y_j^* x_k - q_{kj}x_j y_k^*)\Big\}; \tag{8}$$

hence (A) will follow if it is shown that the above limit, which certainly exists, is zero. To this end, one need note only that, by (6), the right side of (8) becomes

$$l = 2^{-\frac{1}{2}}\lim_{n\to\infty} n^{\frac{1}{2}}(y_{n+1}^* x_n - x_{n+1}y_n^*).$$

Since

$$\sum |y_{n+1}^* x_n - x_{n+1}y_n^*| \leqslant 2(\sum |y_n|^2 \sum |x_n|^2)^{\frac{1}{2}},$$

it readily follows from the divergence of the series $\sum n^{-\frac{1}{2}}$ that $l = 0$, and the proof of (A) is complete. Next, the theorem (I) will be proved.

Suppose that f is any function of L^2 for which QX is in $\mathfrak{H}$, where $X = \{x_k\}$ is defined by (4) for $\phi_k = \xi_k$ and Q is defined by (6). Let N denote an arbitrary positive number and let $g = g_N$ denote any function of L^2 which is zero outside the finite interval $-N \leqslant x \leqslant N$ and which, in addition, satisfies

$$\int_{-\infty}^{\infty} |g|^2\,dx = \int_{-N}^{N} |g|^2\,dx = 1. \tag{9}$$

It is clear that xg is in L^2 and hence that QY is in $\mathfrak{H}$. An application of (A), in conjunction with the Parseval relation, now yields

$$\int_{-N}^{N} xfg^*\,dx = \sum_{k=1}^{\infty}\int_{-\infty}^{\infty} xf\xi_k^*\,dx \int_{-\infty}^{\infty} g^*\xi_k\,dx. \tag{10}$$

(It is to be noted that the functions f, $x\xi_k$ belong to L^2.) An application of the Schwarz inequality, the Parseval relation, and (9) imply

$$\Big|\int_{-N}^{N} xfg^*\,dx\Big|^2 \leqslant \sum_{k=1}^{\infty}\Big|\int_{-\infty}^{\infty} xf\xi_k^*\,dx\Big|^2 < \infty. \tag{11}$$

† A. Wintner, *Spektraltheorie der unendlichen Matrizen* (Leipzig, 1929), 237.

The last inequality follows from the assumption that QX belongs to $\mathfrak{H}$. If g is chosen to be $xf\Big/\Big(\int_{-N}^{N}|xf|^2\,dx\Big)^{\frac{1}{2}}$ on $-N \leqslant x \leqslant N$, and zero elsewhere, relation (11) implies that

$$\int_{-N}^{N}|xf|^2\,dx < \text{constant (independent of } N),$$

for all positive N, and hence that (7) holds. This completes the proof of (I).

Next, a theorem similar in form to (I) will be stated for the momentum matrix P, where

$$P = (P_{jk}) = i\int_{-\infty}^{\infty}\phi_k'\phi_j^*\,dx. \tag{12}$$

(II) *Let f satisfy* (3), *so that f is in L^2 and the corresponding vector X is in $\mathfrak{H}$, and suppose that f has a continuous second derivative on*

$$-\infty < x < \infty.$$

In addition, suppose that X and P are defined by (4) *and* (12), *for $\phi_k = \xi_k$, and that PX belongs to $\mathfrak{H}$. Then f' belongs to L^2 (and hence, as is easily seen from a partial integration of* (13) *below, PX must correspond to if').*

It is clear that all functions ξ_*^k belong to L^2 and that, as a consequence of a partial integration and the Parseval relation, the equation

$$(PX)_k = -i\int_{-\infty}^{\infty}f\xi_k^{*\prime}\,dx \tag{13}$$

is valid. Hence, if g denotes any function with a continuous first derivative on $-\infty < x < \infty$ and which is zero outside a finite interval $-N \leqslant x \leqslant N$, one obtains

$$(X, PY) = i\int_{-N}^{N}f'g^*\,dx \tag{14}$$

and

$$(PX, Y) = -i\sum_k\int_{-N}^{N}g^*\xi_k\,dx\int_{-\infty}^{\infty}f\xi_k^{*\prime}\,dx. \tag{15}$$

It should be noted that, in the derivation of (14), use is made of the assumptions on g together with two partial integrations. As follows readily from the proof of (A), a lemma similar to (A) is valid also for the momentum matrix, in the case of the linear oscillator (P having a form† similar to that of Q in relation (6)), so that the quantities of (14) and

† E. C. Kemble, loc. cit.

(15) are equal; hence, an application of the Schwarz inequality to (14), together with the Parseval relation and the assumption that PX is in $\mathfrak{H}$, yields

$$\left| \int_{-N}^{N} f'g^* \, dx \right|^2 < \text{constant} \int_{-N}^{N} |g|^2 \, dx, \tag{16}$$

where the constant is independent of N. The proof can now be completed in a manner similar to that above. One first defines

$$g(x) = f'(x) \Big/ \left(\int_{-N+\epsilon}^{N-\epsilon} |f'|^2 \, dx \right)^{\frac{1}{2}}$$

on some interval $-N+\epsilon \leqslant x \leqslant N-\epsilon$, where $\epsilon = \epsilon(N)$ is a small positive number such that

$$\left| \left(\int_{-N}^{-N+\epsilon} + \int_{N-\epsilon}^{N} \right) f'g \, dx \right| < \text{constant (independent of } N).$$

Furthermore, it can be arranged that $g(x) \equiv 0$ outside $-N \leqslant x \leqslant N$, that $g(x)$ is defined on the two 'ϵ-intervals' so as to have a continuous derivative on $-\infty < x < \infty$, and that

$$\int_{-N}^{N} |g|^2 \, dx \leqslant \int_{-N+\epsilon}^{N-\epsilon} |g|^2 \, dx + \text{constant } (= 1 + \text{constant}),$$

where the constant is again independent of N. The above requirements are easily fulfilled (the assumptions concerning the derivative of g on $-N+\epsilon \leqslant x \leqslant N-\epsilon$ being a consequence of the assumption that f'' is continuous) and lead easily, via (16), to the conclusion that f' is in L^2. This completes the proof of (II).

In order to prove the remaining assertions made in the introduction, it will be shown by examples that the claims of theorems (I) and (II) can become false if, instead of the particular linear oscillator sequence $\xi_1, \xi_2, \ldots$, an arbitrary orthonormal sequence $\phi_1, \phi_2, \ldots$ is allowed. Consider the half-line $1 \leqslant x < \infty$ and on it the set $\Psi = \{\psi\}$ of all functions ψ which are zero outside a finite interval and which satisfy

$$\int_1^\infty |\psi|^2 \, dx < \infty \quad \text{and} \quad \int_1^\infty \psi \, dx = 0. \tag{17}$$

It will first be shown that the system Ψ is complete on $1 \leqslant x < \infty$. In other words, the only function g on $1 \leqslant x < \infty$ for which

$$\int_1^\infty |g|^2 \, dx < \infty \quad \text{and} \quad \int_1^\infty g\psi^* \, dx = 0 \quad \text{for all } \psi \text{ in } \Psi, \tag{18}$$

is the function $g \equiv 0$. Let c be a constant $\geqslant 1$ and consider the set of functions

$$\psi = \begin{cases} \text{either } \sin k(x-c) \text{ or } \cos k(x-c) & (c \leqslant x \leqslant c+2\pi), \\ 0, \text{ otherwise,} \end{cases}$$

where $k = 1, 2, \ldots$. It is clear that (17) holds and that (18) implies that $g =$ constant on every interval of length 2π on $1 \leqslant x < \infty$. Since g is in L^2, then $g \equiv 0$, as was to be shown.

By extracting functions from Ψ and finite linear combinations of these (which also belong to Ψ) one can obtain an orthonormal, complete sequence $\psi_1, \psi_2, \ldots$, on $1 \leqslant x < \infty$, satisfying (17) for $\psi = \psi_k$. Let the domain of definition of the functions ψ_k be extended by defining $\psi_k \equiv 0$ on $-\infty < x < 1$. Now let $\psi_{-1}, \psi_{-2}, \ldots$ denote an arbitrary orthonormal, complete sequence of functions on $-\infty < x < 1$ and define $\psi_{-k} \equiv 0$ on $1 \leqslant x < \infty$. In addition suppose that (1) holds for $\phi_k = \psi_{-k}$. (That (1) holds for the ψ_k is obvious.) Let now the function f of L^2 be defined to be 0 or $1/x$ according as $-\infty < x < 1$ or $1 \leqslant x < \infty$. Then $xf \equiv 1$ on $1 \leqslant x < \infty$ and, by (17), one has

$$\int_{-\infty}^{\infty} xf\psi_k^* \, dx = \int_{1}^{\infty} \psi_k^* \, dx = 0 \quad \text{(for all } k\text{)}. \tag{19}$$

Since $xf \equiv 0$ on $-\infty < x < 1$, it is clear that

$$\int_{-\infty}^{\infty} xf\psi_{-k}^* \, dx = 0 \quad \text{(for all } k\text{)}. \tag{20}$$

Furthermore, the orthonormal sequence $\{\phi_k\} = \{\psi_k, \psi_{-k}\}$ is complete on $-\infty < x < \infty$, satisfies (1), and, by (19) and (20), is such that the relations

$$\int_{-\infty}^{\infty} xf\phi_k^* \, dx = 0$$

are valid. In other words, $QX = 0$ (the zero vector); hence QX certainly belongs to $\mathfrak{H}$, although xf is not in L^2.

The construction of an orthonormal sequence $\{\phi_k\}$ for which PX is in $\mathfrak{H}$, although f' is not in L^2, is similar but slightly more involved. An outline of the construction is as follows: Let f be in L^2 and have a continuous second derivative on $-\infty < x < \infty$ and be such that $f' \neq 0$ on $1 \leqslant x < \infty$. Furthermore, suppose that

$$\int_{-\infty}^{1} |f'|^2 \, dx < \infty \quad \text{but} \quad \int_{-\infty}^{\infty} |f'|^2 \, dx = \infty. \tag{21}$$

The existence of such a function is clear. Let $\Psi = \{\psi\}$ denote the set of all functions, with continuous first derivatives on $-\infty < x < \infty$, and which, in addition, are zero on $-\infty < x < 1$ and also outside a (sufficiently large) finite interval of $1 \leqslant x < \infty$, and finally, which satisfy

$$\int_{-\infty}^{\infty} f'\psi^* \, dx = 0.$$

(In particular, all functions ψ' are in L^2.) Then, again, the set Ψ is complete on $1 \leqslant x < \infty$. The proof follows easily in a manner similar to that above if one notes that the functions $\sin k(x-c)/f'(x)$, $\cos k(x-c)f'(x)$, if properly 'smoothed out' at the end-points of the intervals

$$c \leqslant x \leqslant c+2\pi,$$

where $c \geqslant 1$, are functions of Ψ. An orthonormal, complete sequence of functions from Ψ can be constructed so that

$$\int_{1}^{\infty} f'\psi_k^* \, dx = 0. \tag{22}$$

Again, let $\{\psi_{-k}\}$ denote an orthonormal complete sequence on

$$-\infty < x < 1$$

and define $\psi_{-k} \equiv 0$ on $1 \leqslant x < \infty$. In addition, it will be supposed that each ψ_{-k} has a continuous derivative ψ'_{-k} on $-\infty < x < \infty$, for which ψ'_{-k} belongs to L^2. Then $\{\phi_k\} = \{\psi_k, \psi_{-k}\}$ is an orthonormal complete sequence on $-\infty < x < \infty$ and

$$\sum_k \left| \int_{-\infty}^{\infty} f'\phi_k^* \, dx \right|^2 < \infty$$

holds; cf. (22) and the first relation of (21). Furthermore, as follows from a partial integration of a relation corresponding to (13), but where ξ_k is replaced by the present ϕ_k, one obtains

$$(PX)_k = i \int_{-\infty}^{\infty} f'\phi_k^* \, dx.$$

Hence PX is in $\mathfrak{H}$ although f' is not in L^2.

Actually, with slight modifications of the above procedure, one can construct a system $\{\phi_k\}$ for which even $PX = 0$, although f' is not in L^2. One should arrange first that $f \equiv 0$ on $-\infty < x < 1$ and then that $\int_1^X |1/f'|^2 \, dx < \infty$, at least for every finite $X \geqslant 1$. (This latter condition was automatically fulfilled above, since there f' was continuous and not

zero on $1 \leqslant x < \infty$.) In addition, as before, f should, and f' should not, belong to L^2. The existence of functions f satisfying the required conditions is clear. It is to be noted that (22) now holds not only for k but also for $-k$ and that consequently $PX = 0$.

Appendix 1. It is well known that each of the operators x and $i\,d/dx$, on $-\infty < x < \infty$, has a continuous spectrum, consisting of all real numbers, and no point spectrum. However, as is seen from the above discussion, it is quite possible that there exist matrix 'representations' of these operators, defined by (2) and (12), for which the number 0 is in the corresponding point spectrum. Thus, for proper corresponding choices of $\{\phi_k\}$, the equations $QX = 0$ and $PX = 0$ can possess solutions $X \neq 0$ in $\mathfrak{H}$.

Appendix 2. It is easy to obtain a certain implication of (I). Consider the case of the linear oscillator, so that $\phi_k = \xi_k$ and Q is defined by (6), and suppose that f is in L^2 (and hence, X is in $\mathfrak{H}$). If, in addition, QX is in $\mathfrak{H}$, then, by (I), xf is in L^2. Another application of (I) shows that, if, in addition, $Q(QX)$ is in $\mathfrak{H}$, then also x^2f is in L^2. In general, if f is in L^2, $X = X_1$ (in $\mathfrak{H}$), and if the vectors $X_{k+1} = QX_k$ are in $\mathfrak{H}$ for $k = 1, 2, \ldots, N$, then the functions $xf, x^2f, \ldots, x^Nf$ are in L^2. It is then an easy consequence of the Schwarz inequality that the relations

$$\int_{-\infty}^{\infty} |x|^k |f|^2\, dx < \infty \quad (k = 1, 2, \ldots, 2N),$$

must hold. Hence, if f denotes any wave function for which all vectors $X_1, \ldots, X_{N+1}$ belong to $\mathfrak{H}$, then the first $2N$ moments of the distribution function σ, with density $|f|^2$, i.e. $d\sigma = |f|^2 dx$, are finite.† (It is to be emphasized that this result has been proved in case the functions ϕ_k are the normalized eigenfunctions of the *linear oscillator* boundary value problem $(\phi_k = \xi_k)$.) Similar results for the momentum matrix P follow from (II).

† Cf., e.g., V. Rojansky, *Introductory Quantum Mechanics* (New York, 1946), 92.

NON-HOMOGENEOUS LATTICES IN THE PLANE

By A. M. MACBEATH (*North Staffordshire University*)

[Received 2 February 1952]

0.1. Introduction

THIS paper falls into three parts. The main object of the paper is to give a new proof of a theorem of the author's (**8**) about the lattice properties of unbounded convex regions in the plane. This proof is contained in Part I. It is much shorter than the other proof and avoids the implicit use of the axiom of choice; but the other proof works in n-dimensional space, whereas this one, in its present form, is applicable only to the plane. Part III of this paper contains an application of the theorem of Part I to the product of two non-homogeneous linear functions; the results are similar to those of Blaney (**1**). In Part II certain critical determinants needed in Part III are evaluated. This evaluation is more or less a mechanical application of Minkowski's method, except that a certain lemma (generalizing a result of Mahler) has to be proved first.

0.2. Notation, definitions, etc.

Generally speaking, the notation is the same as in (**5**), (**8**). Small Roman letters denote points of the plane, small Greek letters denote real numbers, capital letters point-sets. The notation (x_1, x_2) is used for the two coordinates of the point x, and, in the usual vector notation, λx, $x+y$ denote the points with coordinates $(\lambda x_1, \lambda x_2)$, (x_1+y_1, x_2+y_2) respectively. The symbol $|x|$ denotes distance from the origin, and $V(K)$ is used for the area of the point-set K.

Point-sets are combined as in the following example:

$$a+\alpha P-Q$$

denotes the set of all points that can be written in the form

$$a+\alpha x-y \quad (x \in P,\ y \in Q);$$

fr K, int K, $\bar{K}$ denote the frontier, interior, and closure of K respectively. The symbols $\cup$, $\cap$ are employed as usual for the union and intersection of sets.

Quart. J. Math. Oxford (2), 3 (1952), 268–81.

Definition 1. A point-set K is called *symmetrical in* a point u if $K = 2u-K$; u is then called the *centre of symmetry* of K. A non-empty bounded set cannot have more than one centre of symmetry. For if $u \neq v$ and

$$K = 2u-K, \qquad K = 2v-K,$$

we have $K = 2(u-v)+K$, and so by induction $K = 2\xi(u-v)+K$ for every integer ξ. Thus K contains point-pairs whose distance apart is $2\xi|(u-v)|$, contradicting the assumption that K is bounded.

Definition 2. A *lattice* Λ is defined to be a set of points of the form $p+\xi x+\eta y$, where p is any vector, x, y are linearly independent vectors, and ξ, η run over the integers. The determinant $d(\Lambda)$ of Λ is the absolute value of the determinant $x_1 y_2 - x_2 y_1$.

Definition 3. K is *convex* if $K \supset \bigcup_{0<\lambda<1} \{\lambda K+(1-\lambda)K\}$.

K is a *convex body* if K is bounded, closed, convex, and has an interior.

Definition 4. If K is a convex body symmetrical in u, the *critical determinant* $\delta(K)$ is defined by the relation

$$\delta(K) = \inf d(\Lambda) \quad [u \in \Lambda, \Lambda \cap K = \{u\}].$$

We shall require two known properties of $\delta(K)$.

If $u \in \Lambda$, and $d(\Lambda) \leqslant \delta(K)$,

then there is a point of Λ, other than u, in K. (1)

The second is *Hlawka's theorem*:

$$\delta(K) \leqslant V(K). \tag{2}$$

I. The Non-homogeneous Lattice Theorem

I.1. Statement of the result

Let R be a closed convex set containing inner points. Suppose further that R is not the whole plane, a half-plane, or an infinite strip. Then there are two distinct tac-lines of R that are not parallel. Let us call these lines L_1 and L_2. If $u \in R$, the set $R \cap (2u-R)$ is contained in the parallelogram with sides L_1, L_2, $2u-L_1$, $2u-L_2$. Thus $R \cap (2u-R)$ is always a convex body and the numbers $V(R \cap (2u-R))$, $\delta(R \cap (2u-R))$ are finite. Henceforth $V(R \cap (2u-R))$ will be denoted by $f(u)$.

Theorem 1. *Let R be a closed convex set satisfying the conditions stated above. Let Λ be a lattice. If $\Lambda \cap R \neq 0$, then there is a point $x \in \Lambda \cap R$ such that*

$$\delta(R \cap (2x-R)) < d(\Lambda).$$

Since δ is very closely related to the volume V [see (2)], it is natural that the properties of $f(u)$ should play an important part in the proof. The main result in this connexion is of some interest in itself.

THEOREM 2. *Hypothesis*: $u \in \operatorname{int} R$; $u+v \in \operatorname{fr} R$; $f(u+\alpha v) \leqslant f(u)$ *for each* α *such that* $0 \leqslant \alpha \leqslant 1$.

Conclusion: $f(u+\alpha v) \leqslant (1-\alpha^2)f(u)$ *if* $0 \leqslant \alpha \leqslant 1$.

The proof of Theorem 2 depends on some inequalities about convex sets due to Bonnesen [cf. Bonnesen (**2**), Henstock–Macbeath (**5**)]. Only very special cases of these inequalities are needed, and, since the proof is simple, it seemed worth while to establish them from the beginning.

I.2. Bonnesen's lemma

Let K, K' be two convex bodies, with projections of length α, α' on a given line l. We assume without loss of generality that l is the axis of x_1. Let $(\beta, \beta+\alpha)$ be the projection of K and $(\beta', \beta'+\alpha')$ be that of K'. Let $\phi(\tau)$ be the length of the segment in which K is cut by the line $x_1 = \beta+\alpha\tau$, with a corresponding definition for $\phi'(\tau)$.

The segment in which the body $K+K'$ is cut by the line

$$x_1 = \beta+\beta'+(\alpha+\alpha')\tau$$

has then a length at least $\phi(\tau)+\phi'(\tau)$. Hence

$$V(K+K') \geqslant (\alpha+\alpha')\int_0^1 [\phi(\tau)+\phi'(\tau)]\,d\tau = (\alpha+\alpha')\left[\frac{1}{\alpha}V(K)+\frac{1}{\alpha'}V(K')\right].$$

If λ, λ' are positive, we can apply this formula to the bodies λK, $\lambda' K'$. Since the projection of λK has length $\lambda\alpha$, and the area of λK is $\lambda^2 V(K)$, we have

$$V(\lambda K+\lambda' K') \geqslant (\lambda\alpha+\lambda'\alpha')\left[\frac{\lambda}{\alpha}V(K)+\frac{\lambda'}{\alpha'}V(K')\right]. \tag{3}$$

From this relation it follows by Cauchy's inequality that

$$V(\lambda K+\lambda' K')^{\frac{1}{2}} \geqslant \lambda V(K)^{\frac{1}{2}}+\lambda' V(K')^{\frac{1}{2}}. \tag{4}$$

This is the two-dimensional case of the well-known Brunn–Minkowski theorem.

Each line parallel to the x_1-axis cuts K in a segment. If α denotes the length of one of the longest of these segments, and α' is similarly defined in relation to K', the formula (3) will again hold. This follows by symmetrizing in the x_2-axis and then applying the previous result [see Bonnesen (**2**), Bonnesen–Fenchel (**3**, 94–6)]. Suppose, in particular, that K, K' have centres of symmetry on the axis of x_1. Then no line

parallel to the axis of x_1 will have a longer section with K (or K') than the axis itself has. Hence we have

LEMMA 1. *If K, K' have centres of symmetry on a line l and if α, α' are the lengths of $l \cap K$, $l \cap K'$ respectively; then the formula* (3) *holds.*

Definition. Let $\{K_\alpha\}$ be a set of convex bodies, associated with the values of a real parameter α. $\{K_\alpha\}$ is called a *concave system* of bodies if, whenever

$$\gamma = \lambda\alpha + (1-\lambda)\beta \quad (0 < \lambda < 1),$$

then

$$K_\gamma \supset \lambda K_\alpha + (1-\lambda)K_\beta.$$

If $\{K_\alpha\}$ is a concave system, it follows from (4) that the square root of $V(K_\alpha)$ is a concave function of α, and hence that $V(K_\alpha)$ has left and right derivatives at every point [see Hardy, Littlewood, and Pólya (**4**)]. The symbols D_- and D_+ are used to denote left and right derivatives respectively.

LEMMA 2. *Suppose that $\{K_\alpha\}$ is a concave system of bodies, each having a centre of symmetry on the same line l. Set $V_\alpha = V(K_\alpha)$ and let l_α be the length of the segment $l \cap K_\alpha$. Then, if $\beta > \alpha$,*

$$(\beta-\alpha)D_+V_\alpha \geqslant (-2l_\alpha + l_\beta)\frac{V_\alpha}{l_\alpha} + \frac{l_\alpha}{l_\beta}V_\beta.$$

Proof. If $0 < \lambda < 1$, we have, from Lemma 1 with $\lambda' = 1-\lambda$,

$$V(\alpha+\lambda(\beta-\alpha)) \geqslant ((1-\lambda)l_\alpha + \lambda l_\beta)\left[(1-\lambda)\frac{V_\alpha}{l_\alpha} + \lambda\frac{V_\beta}{l_\beta}\right].$$

Hence

$$\frac{V(\alpha+\lambda(\beta-\alpha))-V_\alpha}{\lambda} \geqslant (-2l_\alpha + l_\beta)\frac{V_\alpha}{l_\alpha} + \frac{l_\alpha}{l_\beta}V_\beta + O(\lambda).$$

By letting $\lambda \to 0$, the result follows.

COROLLARY. *If $\alpha > \beta$, the same formula holds with the left-hand side replaced by $(\beta-\alpha)D_-V_\alpha$. This result is obtained by applying Lemma 2 to the concave system $\{L_\alpha\}$ of bodies defined by $L_\alpha = K_{-\alpha}$.*

LEMMA 3. *Under the same assumptions as Lemma 2, we have*

$$D_-V_\alpha \geqslant D_+V_\alpha.$$

Proof. $V_\alpha^{\frac{1}{2}}$ is concave, so by (**4**) [91, Theorem 111],

$$D_-(V_\alpha)^{\frac{1}{2}} \geqslant D_+(V_\alpha)^{\frac{1}{2}}.$$

Since $D_\pm V_\alpha = 2V_\alpha^{\frac{1}{2}} D_\pm(V_\alpha)^{\frac{1}{2}}$, the result follows.

I.3. Proof of Theorem 2

Let $K_\alpha = R \cap (2(u+\alpha v)-R)$. I shall show that $\{K_\alpha\}$ is a concave system of bodies. Since R is convex, we have

$$R \supset \lambda R+(1-\lambda)R \quad \text{if } 0 < \lambda < 1. \tag{5}$$

Now, if $\gamma = \lambda\alpha+(1-\lambda)\beta$, it follows from (5) that

$$2(u+\gamma v)-R \supset \lambda(2(u+\alpha v)-R)+(1-\lambda)(2(u+\beta v)-R). \tag{6}$$

Hence, from (5), (6), by taking intersections,

$$K_\gamma \supset \lambda K_\alpha+(1-\lambda)K_\beta.$$

Let l be the line consisting of all the points that can be written in the form $u+\xi v$, where ξ varies. Let $R \cap l$ be given by the relation

$$-\tau \leqslant \xi \leqslant 1,$$

where τ may possibly be infinite. Since $u \in \operatorname{int} R$, τ must be positive. As in Lemma 2, let l_α be the length of $l \cap K_\alpha$. Now $l \cap K_\alpha$ is given by the two relations

$$-\tau \leqslant \xi \leqslant 1, \qquad 2\alpha-1 \leqslant \xi \leqslant 2\alpha+\tau. \tag{7}$$

If we set $\sigma = \frac{1}{2}(1-\tau)$, there are two cases to consider.

Case 1: $\alpha \geqslant \sigma$. The relations (7) reduce to

$$2\alpha-1 \leqslant \xi \leqslant 1,$$

and we have $l_\alpha = 2(1-\alpha)|\xi|$.

Case 2: $\alpha \leqslant \sigma$. The relations (7) reduce to

$$-\tau \leqslant \xi \leqslant 2\alpha+\tau,$$

and we have $l_\alpha = 2(\alpha+\tau)|\xi| = 2(1-2\sigma+\alpha)|\xi|$.

The two different expressions just found for l_α make it necessary to split up the proof of Theorem 2 to cover the following three cases.

Case 1: $\sigma \leqslant 0$. Then $l_\alpha = 2(1-\alpha)|\xi|$ for $0 \leqslant \alpha \leqslant 1$. We have, by Lemma 2,

$$\alpha D_+ V_0 \geqslant (-1-\alpha)V_0+\frac{1}{1-\alpha}V_\alpha.$$

Since $V_0 \geqslant V_\alpha$ for all positive α, $D_+ V_0 \leqslant 0$, so

$$V_\alpha \leqslant (1-\alpha^2)V_0.$$

Case 2: $\sigma > 0$; $0 < \alpha \leqslant \sigma$. Again apply Lemma 2. We have

$$0 \geqslant D_+ V_0 \geqslant (-\tau+\alpha)\frac{V_0}{\tau}+\frac{\tau}{\tau+\alpha}V_\alpha,$$

$$V_\alpha \leqslant \left(1-\frac{\alpha^2}{\tau^2}\right)V_0 < (1-\alpha^2)V_0.$$

From the first of these inequalities, since $V_\sigma \geqslant 0$, we have $\sigma \leqslant \tau$ and so, since $\sigma = \frac{1}{2}(1-\tau)$, we have

$$\sigma \leqslant \tfrac{1}{3}. \tag{8}$$

Case 3: $\alpha > \sigma > 0$. Apply Lemma 2 three times to obtain the relations:

$$0 \geqslant \frac{-1+3\sigma}{1-2\sigma} V_0 + \frac{1-2\sigma}{1-\sigma} V_\sigma,$$

$$-\sigma D_- V_\sigma \geqslant -\frac{V_\sigma}{1-\sigma} + \frac{1-\sigma}{1-2\sigma} V_0,$$

$$(\alpha-\sigma) D_+ V_\sigma \geqslant -\frac{1-2\sigma+\alpha}{1-\sigma} V_\sigma + \frac{1-\sigma}{1-\alpha} V_\alpha.$$

By a straightforward elimination process, using Lemma 3 to get rid of the derivatives, and using the relation (8), we find

$$\frac{V_\alpha}{(1-\alpha^2)V_0} \leqslant \left\{1-5\sigma+\frac{8\sigma^2}{1+\alpha}\right\} D^{-1}, \tag{9}$$

where $D = 1-5\sigma+8\sigma^2-4\sigma^3 = (1-\sigma)(1-2\sigma)^2 > 0$, by (8). We have, since $\sigma < \alpha \leqslant 1$,

$$\frac{8\sigma^2}{1+\alpha} < \frac{8\sigma^2}{1+\sigma} = 8\sigma^2 - \frac{8\sigma^3}{1+\sigma} < 8\sigma^2 - 4\sigma^3.$$

Hence the numerator of (9) is less than the denominator, and Theorem 2 is established.

I.4. Proof of Theorem 1

LEMMA 4. *If $u \in \operatorname{int} R$, and $u+\alpha v \in R$ for all positive α, then $f(u+\alpha v)$ tends to infinity with α.*

Proof. Let w be a vector perpendicular to v. Then, since $u \in \operatorname{int} R$, there is a positive number ϵ such that $u \pm \epsilon w \in R$. Let T denote the triangle with vertices $u \pm \epsilon w$, $u+3\alpha v$. Since R is convex, $R \supset T$, and so

$$R \cap (2(u+\alpha v) - R) \supset T \cap (2(u+\alpha v) - T).$$

The area of the expression on the right is $2\alpha|v|\epsilon|w|$, which tends to infinity with α.

LEMMA 5. *Let $u \in \operatorname{int} R$, $w \neq 0$. Then one or other of the following two statements is true:*

(i) $f(u) \geqslant f(u+\alpha w)$ *for all* $\alpha > 0$,

(ii) $f(u) \geqslant f(u-\beta w)$ *for all* $\beta > 0$.

Proof. Suppose the contrary,

i.e. $f(u+\alpha w) > f(u), \quad f(u-\beta w) > f(u) \quad (\alpha, \beta > 0).$

Apply (4) with $\lambda = \dfrac{\beta}{\alpha+\beta}, \quad \lambda' = \dfrac{\alpha}{\alpha+\beta}.$

Then $$f(u)^{\frac{1}{2}} \geqslant \frac{\beta}{\alpha+\beta} f(u+\alpha w)^{\frac{1}{2}} + \frac{\alpha}{\alpha+\beta} f(u-\beta w)^{\frac{1}{2}}$$

$$> f(u)^{\frac{1}{2}}, \text{ a contradiction.}$$

LEMMA 6. *If $u \in \Lambda \cap \operatorname{int} R$, and $\delta(R \cap (2u-R)) \geqslant d(\Lambda)$, then there is a point $v \in \Lambda \cap R$ such that*

$$f(v) \leqslant \tfrac{3}{4} f(u).$$

Proof. Since $\delta(R \cap (2u-R)) \geqslant d(\Lambda)$, $u \in \Lambda$, there is, by (1), a point $u+w \in \Lambda \cap R \cap (2u-R)$ $(w \neq 0)$. Now, if $u+w \in \Lambda$, so also is $u-w$. Hence, after Lemma 5, we may assume, on replacing w by $-w$ if necessary, that $f(u+\alpha w) \leqslant f(u)$ for all $\alpha > 0$. Then, by Lemma 4, there are positive values of α such that $u+\alpha w$ does not lie in R. Let β be the lower bound of such α. Then $u+\beta w \in \operatorname{fr} R$. Since $u+w \in R$, β is not less than 1. Hence

$$2[\beta] \geqslant [\beta]+1 \geqslant \beta. \tag{10}$$

Let $v = u+[\beta]w$. We have $v \in \Lambda$, and, by Theorem 2,

$$f(v) \leqslant \left\{1 - \left(\frac{[\beta]}{\beta}\right)^2\right\} f(u) \leqslant \tfrac{3}{4} f(u), \quad \text{by (10).}$$

This proves Lemma 6.

It is now simple to complete the proof of Theorem 1. By hypothesis, there is a lattice-point, say u_0, $\in \operatorname{int} R$. If $\delta(R \cap (2u_0-R)) < d(\Lambda)$, Theorem 1 holds. Suppose, then, that $\delta(R \cap (2u_0-R)) \geqslant d(\Lambda)$. By Lemma 6, there is then a point $u_1 \in \Lambda \cap R$, such that $f(u_1) \leqslant \frac{3}{4} f(u_0)$. If, again, $\delta(R \cap (2u_1-R)) \geqslant d(\Lambda)$, the process can be repeated, to obtain successively $u_2, u_3, \ldots$, where

$$f(u_n) \leqslant (\tfrac{3}{4})^n f(u_0).$$

Since, by Hlawka's theorem (2), $\delta(R \cap (2u-R)) \leqslant f(u)$, the inequality $\delta(R \cap (2u_n-R)) < d(\Lambda)$ must hold for some u_n of the sequence. This completes the proof of Theorem 1.

II. A Certain Critical Determinant

Before Theorem 1 can be applied to the product of two non-homogeneous linear functions, the value of a certain critical determinant must be known. A generalization of a lemma of Mahler's is needed in order to derive the result.

II.1. Mahler's Lemma

Definition 5. Let K be a convex body symmetrical about the origin. Let Λ be a lattice containing the origin. (In Part II all lattices will be assumed to contain the origin.) Then Λ is called *geometrically critical* if Λ has

(i) no point except 0 in the interior of K,

(ii) six points $\pm p_1$, $\pm p_2$, $\pm(p_1+p_2)$ on the boundary of K.

It has been shown by Minkowski that a critical lattice is always geometrically critical, but of course the converse does not hold. The problem of finding the critical determinant of K is thus reduced to that of finding the geometrically critical lattice of minimum determinant.

Definition 6. If K possesses the property that, if $0 < \lambda < 1$,

$$p \in \overline{K},\ q \in \overline{K} \Rightarrow (1-\lambda)p+\lambda q \in \operatorname{int} K,$$

then K is called *strictly convex*.

Definition 7. If p, p' are two linearly independent boundary-points of K (i.e. $p \neq \pm p'$), then every point of the plane can be written in the form $\lambda p+\mu p'$. If q is a boundary-point, and $q = \lambda p+\mu p'$, where λ, μ are non-negative, we shall say that q *lies between* p and p'.

Now consider a geometrically critical lattice Λ. Let p_1, p_2 be as in Definition 1. Rename the points of Λ that lie on the boundary of K as follows:

$$\begin{array}{ccc} p_1 = q_1, & p_1+p_2 = q_2, & p_2 = q_3, \\ -p_1 = q_4, & -p_1-p_2 = q_5, & -p_2 = q_6. \end{array}$$

The boundary of K is then divided into six sets, defined as follows:

C_i consists of the points lying between q_i and q_{i+1} $(i = 1,..., 5)$,

C_6 consists of the points lying between q_6 and q_1.

It is easily verified that these sets are disjoint, apart from the end-points q_i.

LEMMA 7. [See Mahler (**6**), Lemma 6, where the result is proved for critical, instead of geometrically critical, lattices.] *Let K be strictly convex. Let Λ_0 be any geometrically critical lattice dividing the boundary of K into the six arcs $C_1,..., C_6$. Let Λ_1 be a geometrically critical lattice which does not coincide with Λ_0. Then there is exactly one point of Λ_1 in each of the six arcs C.*

Proof. Since Λ_1 is geometrically critical, there are six points of Λ_1 on $\operatorname{fr} K$. It suffices, then, to show that no two of them can lie in the same arc, say C_1.

Suppose, then, that, if possible, $q = \lambda q_1 + \mu q_2$, $q' = \lambda' q_1 + \mu' q_2$ are lattice-points on the boundary of K, and that λ, μ, λ', μ' are non-negative. I shall show that, unless q is q_1 or q_2, the following inequalities hold:

$$0 < \lambda < 1, \qquad 0 < \mu < 1, \qquad \lambda + \mu > 1. \tag{11}$$

Suppose first, if possible, that $\lambda + \mu \leqslant 1$. Then

$$q = \lambda q_1 + \mu q_2 + (1 - \lambda - \mu)0 \in \operatorname{int} K,$$

by strict convexity, unless two of the three coefficients are zero, i.e. unless $q = q_1$ or q_2. This contradiction proves the last inequality.

Suppose again that $\lambda \geqslant 1$. Since $q_1 - q_2 = -p_2 \in \operatorname{fr} K$, we have

$$q_1 = \frac{1}{\lambda+\mu} q + \frac{\mu}{\lambda+\mu}(q_1 - q_2) + \frac{\lambda-1}{\lambda+\mu} 0 \in \operatorname{int} K,$$

unless $q = q_1$.

Now we can suppose that q, q' are not the points q_1, q_2 in some order, for then the lattice Λ_1 would coincide with Λ_0, contrary to hypothesis. Then, by what we have shown, at least one of the pairs (λ, μ), (λ', μ') will satisfy (11), and the other will satisfy (11) if we replace the strict inequality signs by weak ones. Since (11) implies $1 < \lambda + \mu < 2$, we derive

$$|\lambda - \lambda'| < 1, \qquad |\mu - \mu'| < 1, \qquad |(\lambda - \lambda') + (\mu - \mu')| < 1.$$

These inequalities mean that $q - q'$ $(\in \Lambda_1)$ lies in the interior of the convex cover of $q_1, \ldots, q_6$ and hence in the interior of K. This contradiction proves Lemma 7.

II.2. Application

Mahler's lemma will be used in the proof of the following result.

THEOREM 3. *If* $0 \leqslant \lambda < 1$, *the critical determinant of the set*

$$x^2 - \lambda y^2 + (\lambda + 1)|y| \leqslant 1 \qquad (|y| \leqslant 1) \tag{12}$$

is $\frac{1}{2}\sqrt{(2-\lambda)}$.

Note. Without risk of confusion, I use the notation (x, y) for coordinates instead of that of § 0.2. I shall use this more ordinary notation from now on, since we shall be dealing with special regions. The proof of Theorem 3 is given only for the case $0 < \lambda < 1$. A few simple modifications, left to the reader, will deal with $\lambda = 0$. The result is also true for $\lambda = 1$, when the set becomes a square.

Proof. The set (2) is symmetrical, since the defining relation is unaltered by the substitution $(x, y) \to (-x, -y)$. It is also strictly convex, as can be seen by writing the first inequality in the form

$$\lambda |y|^2 - (\lambda + 1)|y| + (1 - x^2) \geqslant 0.$$

The value of $|y|$ must therefore be either greater than the large zero or less than the small zero of the polynomial on the left. Since the sum of the zeros is $1+\lambda^{-1} > 2$, the second relation $|y| \leqslant 1$ implies that $|y|$ must be less than the small zero, i.e.

$$|y| \leqslant \frac{1}{2\lambda}[(\lambda+1)-\{(\lambda-1)^2+4\lambda x^2\}^{\frac{1}{2}}] = \phi(x), \text{ say.}$$

Since $\phi''(x) < 0$, the strict convexity is proved.

The boundary of (2) consists of two arcs, which I shall call the *upper arc* ($y \geqslant 0$) and the *lower arc* ($y \leqslant 0$).

Let Λ_0 be the geometrically critical lattice generated by the points $p_0\,(1, 0)$, $p_1\,(\frac{1}{2}, \alpha)$, where α is the smaller root of the equation

$$\lambda y^2-(\lambda+1)y+\tfrac{3}{4} = 0.$$

The lattice Λ_0 also contains the boundary-points $p_2\ (-\frac{1}{2}, \alpha)$ and $p_3\ (-1, 0)$. The points p all lie on the upper arc.

If Λ is any geometrically critical lattice, then, by Mahler's lemma, it will contain points q_1, q_2, q_3 on the upper arc, where q_1 lies in the arc p_0p_1, q_2 in the arc p_1p_2, q_3 in the arc p_2p_3. Further $q_2 = q_1+q_3$. (Consideration of the magnitude of the y-coordinate rules out such possibilities as $q_1 = -q_2, q_2 = q_1-q_3$.) Let q_1 be (ξ, η), where, since q_1 lies in the arc p_0p_1, we have

$$\xi^2-\lambda\eta^2+(\lambda+1)\eta = 1, \qquad \tfrac{1}{2} \leqslant \xi \leqslant 1. \tag{13}$$

Let q_2q_3 meet the axis of x at $(s, 0)$. Since q_2q_3 is parallel to Oq, a general point on the line q_2q_3 is given in terms of a parameter t by $(s+t\xi, t\eta)$. The values of t at the points q_2, q_3 where the line cuts the upper arc are given by a quadratic equation obtained by substituting $x = s+t\xi$, $y = t\eta$ in the equation of the upper arc. Since q_2q_3 is equal, as well as parallel, to Oq_1, the difference of the roots is unity. On deriving the condition for this, and using the relation (13) to get rid of all powers of ξ higher than the first, we obtain

$$g(\Delta) = 4\lambda\Delta^2-4\Delta\xi(\lambda+1)-2(\lambda+1)\eta+3 = 0, \tag{14}$$

where $\Delta = -s\eta$ is the determinant of Λ. We shall see that both the roots of (14) are positive, so that the positive value of Δ really is given by $-s\eta$, not $+s\eta$.

The proof of the Theorem is now completed in two stages. Write $\theta = \frac{1}{2}\sqrt{(2-\lambda)}$. Then we have

LEMMA 8. *θ is less than the larger zero of $g(x)$.*

Proof. $$(\lambda-1)^2(4\lambda+1) > 0.$$

Hence $$4\lambda^2(2-\lambda) < (\lambda+1)^2.$$

Take square roots and divide by 4λ. Then

$$\theta < \frac{1}{4}\left(1+\frac{1}{\lambda}\right) \leqslant \tfrac{1}{2}\xi\left(1+\frac{1}{\lambda}\right).$$

Thus θ is less than the mean value of the zeros of $g(x)$.

LEMMA 9. *θ is less than or equal to the smaller zero of $g(x)$.*

Proof. Since $g(x)$ can be positive only if x is greater than both roots or less than both roots, it suffices, after Lemma 8, to prove that $g(\theta) \geqslant 0$. Now $g(\theta) = (\lambda+1)\{3-\lambda-2\eta-2\xi\sqrt{(2-\lambda)}\}$; so it suffices to show that

$$3-\lambda-2\eta \geqslant 2\xi\sqrt{(2-\lambda)}. \tag{15}$$

Neither λ nor η is greater than unity, so the left side is positive, and (15) follows from the same inequality with both sides squared, i.e.

$$(3-\lambda-2\eta)^2 \geqslant 4\xi^2(2-\lambda).$$

Substitution of the value of ξ^2 from (13) reduces this to

$$(\lambda-1)^2(2\eta-1)^2 \geqslant 0,$$

an inequality which is true. This proves Lemma 9.

From Lemma 9 we deduce that $\Delta > \theta$, except possibly when $\eta = \frac{1}{2}$. Corresponding to $\eta = \frac{1}{2}$, we have the geometrically critical lattice generated by the vectors $(\theta, \pm\frac{1}{2})$, which does indeed have determinant θ. Thus θ is the critical determinant, and the theorem is proved.

III. THE PRODUCT OF TWO NON-HOMOGENEOUS LINEAR FUNCTIONS

Let $a > 0$. Let R be the region defined by the inequalities

$$x \geqslant 0, \qquad y \geqslant 0, \qquad xy \geqslant a.$$

It is well known that R is convex and that R contains a point of every lattice [see, e.g., (8)]. If p is the point (u, v), then $R \cap (2p-R)$ is given by the relations

$$0 \leqslant x \leqslant 2u, \qquad 0 \leqslant y \leqslant 2v, \qquad xy \geqslant a, \qquad (2u-x)(2v-y) \geqslant a.$$

Apply the affine transformations:

$$x = u\left\{1-x'\sqrt{\left(1-\frac{a}{uv}\right)}-y'\left(1-\sqrt{\frac{a}{uv}}\right)\right\},$$

$$y = v\left\{1+x'\sqrt{\left(1-\frac{a}{uv}\right)}-y'\left(1-\sqrt{\frac{a}{uv}}\right)\right\}.$$

This maps $R \cap (2p-R)$ on the region of Theorem 3, with

$$\lambda = \left(1-\sqrt{\frac{a}{uv}}\right)\Big/\left(1+\sqrt{\frac{a}{uv}}\right).$$

Since, for any affine transformation T, $\delta(T(K)) = |\det T|\delta(K)$, we find that $\delta(R \cap (2p-R)) = \{\sqrt{(uv)}+3\sqrt{a}\}^{\frac{1}{2}}\{\sqrt{(uv)}-\sqrt{a}\}^{\frac{1}{2}}$. From Theorem 1 we deduce at once that

THEOREM 4a. *There is a point of Λ satisfying the inequalities*

$$xy \geqslant a, \qquad x \geqslant 0, \qquad y \geqslant 0, \qquad \{\sqrt{(xy)}+3\sqrt{a}\}^{\frac{1}{2}}\{\sqrt{(xy)}-\sqrt{a}\}^{\frac{1}{2}} < d(\Lambda).$$

This result may also be expressed in the form

THEOREM 4b. *If $k > 2$ and $d(\Lambda) = a(k-2)(k^2-4)^{\frac{1}{2}}$, then there is a point of Λ in the set*

$$a \leqslant xy < (k-1)^2 a, \qquad x \geqslant 0, \qquad y \geqslant 0.$$

The result in this form is best-possible if k is an integer greater than 2. It suffices to prove this assertion for the particular value $a = (k-2)^{-1}$; for affine maps of the form $(x, y) \to (\alpha x, \alpha y)$ will then prove the result for any positive a.

It suffices, then, to produce a lattice of determinant $(k^2-4)^{\frac{1}{2}}$ with no point in the set

$$\frac{1}{k-2} < xy < \frac{(k-1)^2}{k-2} \quad (x \geqslant 0;\ y \geqslant 0).$$

This is done as follows. Let p, q run over all real numbers that are congruent to $(k-2)^{-1} \pmod 1$. Let α, α^{-1} be the large and small roots of the equation

$$t^2-kt+1 = 0.$$

Then the non-homogeneous lattice given by

$$x = \alpha q - p, \qquad y = p - \alpha^{-1} q$$

has the desired property. To prove this it must be shown that the following relations cannot hold simultaneously:

$$\frac{1}{k-2} < kpq-p^2-q^2 < \frac{(k-1)^2}{k-2}, \tag{16}$$

$$p, q \equiv \frac{1}{k-2} \pmod 1,$$

$$\alpha p - q \geqslant 0, \qquad \alpha q - p \geqslant 0. \tag{17}$$

Now the form $kpq-p^2-q^2$ admits the following linear transformation into itself

$$p' = kp-q, \qquad q' = p. \tag{18}$$

This transformation preserves the property $p \equiv q \equiv (k-2)^{-1} \pmod 1$, and, if we write $\tau = p/q$, τ is transformed by the rule

$$\tau' = k - \tau^{-1}. \tag{19}$$

If there were a pair satisfying (16), (17), we could assume, interchanging p, q if necessary, that the corresponding τ is greater than 1. In the interval $1 \leqslant \tau < \alpha$ the transformation (19) preserves order, i.e.

$$\tau_1 < \tau_2 \Rightarrow \tau_1' < \tau_2'.$$

Consider the sequence

$$\sigma_0 = 1, \quad \sigma_1 = k-1; \quad \ldots, \quad \sigma_i = k-\sigma_{i-1}^{-1}, \quad \ldots.$$

This is a monotone increasing sequence tending to the limit α; so, if a number τ lies between 1 and α, then it falls between σ_μ and $\sigma_{\mu+1}$ for some μ. Hence, if we apply the inverse map of (19) μ times, we obtain a number τ_0 such that $1 \leqslant \tau_0 < k-1$. Going back to (18), this means that, if there is a pair satisfying (16), (17), then there is also a pair satisfying (16), (17) and satisfying in addition the further relation

$$1 \leqslant p/q < k-1. \tag{20}$$

We have only to show that this is impossible. (17) and (20) together imply that p, q are non-negative; for (20) gives p/q positive and the two inequalities of (17), added together, give $p+q \geqslant 0$. Now consider the behaviour of the function $kpq-p^2-q^2$ for fixed q. It has its minima for the interval $q \leqslant p \leqslant (k-1)q$ at the end-points, for its only absolute turning-point is a maximum. Hence

$$kpq-p^2-q^2 \geqslant (k-2)q^2;$$

and the inequality (16) implies

$$(k-2)q^2 < \frac{(k-1)^2}{k-2},$$

so that

$$q < \frac{k-1}{k-2}.$$

The only possibility then remaining for q is $q = (k-2)^{-1}$. Then (20) leaves only two possible values for p:

$$p = \frac{k-1}{k-2}, \qquad p = \frac{1}{k-2}.$$

Neither of these satisfies (16) with strict inequality, and so the result is proved.

Note. The forms $kpq-p^2-q^2$ have occurred before in similar connexions in the work of Segre (**9**) and Blaney (**1**). The above seems to be a new property, since, if k is an integer not less than 4, the result ceases to be best-possible if the conditions $x \geqslant 0$, $y \geqslant 0$ are removed.

The lemma of Part II and the theorem of Part I suffice also to prove another known result. Let R be the region $y \geqslant x^2$ and let p be the point (u, v). $R \cap (2p-R)$ can be reduced by an affine transformation of determinant unity to the set

$$|y|+x^2 \leqslant \alpha,$$

where $\alpha = u-v^2$. The further transformation $(x, y) \to (\alpha^{\frac{1}{2}}x, \alpha y)$ maps this on the set of Theorem 3 with $\lambda = 0$. Hence

$$\delta(R \cap (2p-R)) = \frac{1}{\sqrt{2}}(u-v^2)^{\frac{1}{2}}.$$

Applying Theorem 1, we deduce

THEOREM 5. *There is a point of* Λ *in the set*

$$0 \leqslant y-x^2 < \{2d(\Lambda)^2\}^{\frac{1}{2}}.$$

This result, though best-possible, is merely the first of a series of minima [see (7)], whose existence has not so far been established by geometrical methods.

REFERENCES

1. H. Blaney, 'Indefinite quadratic forms in n variables', *J. of London Math. Soc.* (2) 23 (1948), 153–60.
2. T. Bonnesen, *Les Problèmes des isopérimètres et des isépiphanes* (Paris, 1929).
3. T. Bonnesen und W. Fenchel, 'Theorie der konvexen Körper' (*Ergebnisse der Math.*) (Berlin, 1934).
4. G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities* (Cambridge, 1934).
5. R. Henstock and A. M. Macbeath, 'On the measure of sum-sets (I): The theorem of Brunn, Minkowski, and Lusternik' (in course of publication in *Proc. Lond. Math. Soc.*)
6. K. Mahler, *Proc. Dutch Academy* (Amsterdam) 50 (1947), 98–119.
7. A. M. Macbeath, 'A new sequence of minima in the Geometry of Numbers', *Proc. Cambridge Phil. Soc.* 47 (1951), 266–73.
8. —— 'A theorem on non-homogeneous lattices' (in course of publication in the *Annals of Math.*).
9. B. Segre, 'Lattice-points in infinite domains and asymmetric Diophantine approximation', *Duke Math. J.* 12 (1945), 337–65.

ON AN UNSOLVED PROBLEM IN THE ANALYTIC THEORY OF NUMBERS

By A. Y. FAWAZ (*Cairo*)

[Received 19 February 1952]

1. LET $$\frac{\zeta(2s)}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\lambda(n)}{n^s} \quad (s = \sigma + ti;\ \sigma > 1),$$

$$L(x) = \sum_{n \leqslant x} \lambda(n), \qquad L_0(x) = \sum_{n \leqslant x}{}' \lambda(n) \quad (x > 0).$$

It was conjectured by Pólya (**7**) in 1919 that

$$L(x) \leqslant 0 \quad (x \geqslant 2). \tag{1}$$

The calculations of Pólya confirm this up to $x = 1{,}500$, and the verification has since been extended to $x = 20{,}000$. [See Gupta (**2**), (**3**).] In 1942 Ingham (**4**) gave reasons for supposing that, on the contrary,

$$\lim \begin{matrix} \sup \\ \inf \end{matrix} x^{-\frac{1}{4}} L(x) = \pm\infty \quad (x \to \infty). \tag{2}$$

In a recent paper [Fawaz (**1**)] the explicit formula for $L(x)$ was investigated; in particular, an explicit representation of a closely related integral $I(x)$ as an infinite series was obtained. This was undertaken in the hope that it might throw some light on the conflicting hypotheses of Pólya and Ingham. In this paper I propose to follow out in detail the relationship, suggested in the concluding section of (**1**), between the behaviour of $L(x)$ for large x and of $I(x)$ for small positive x.

I assume the Riemann hypothesis concerning the zeros of $\zeta(s)$, and the subsidiary hypothesis that all the (complex) zeros are simple. This assumption does not affect any of the proposed applications. The explicit formula for $L(x)$ (or $L_0(x)$) is then [see Fawaz (**1**)]

$$L_0(x) = \frac{x^{\frac{1}{2}}}{\zeta(\frac{1}{2})} + \sum_{\rho} \frac{\zeta(2\rho)}{\rho\zeta'(\rho)} x^{\rho} + I(x) \quad (x > 0), \tag{3}$$

where ρ runs through the complex zeros of $\zeta(s)$, the terms of the series are suitably grouped, and

$$I(x) = \frac{1}{2\pi i} \int_{(a)} \frac{\zeta(2s)}{s\zeta(s)} x^s \, ds \quad (0 < a < \tfrac{1}{2}),$$

Quart. J. Math. Oxford (2), 3 (1952), 282–95.

this integral being absolutely convergent for $x > 0$ [(1) § 3]. We may write (3) in the form

$$F_0(u) = \frac{L_0(e^u) - I(e^u)}{e^{\frac{1}{2}u}} = \alpha_0 + \sum_{n=1}^{\infty} (\alpha_n e^{\gamma_n ui} + \bar{\alpha}_n e^{-\gamma_n ui}), \tag{4}$$

where $x = e^u$, u is real, and

$$\alpha_0 = \frac{1}{\zeta(\frac{1}{2})}, \qquad \alpha_n = \frac{\zeta(2\rho)}{\rho\zeta'(\rho)}, \qquad \rho = \rho_n = \tfrac{1}{2} + \gamma_n i,$$

where $n = \pm 1, \pm 2, \pm 3, \ldots$; γ_n is real; $\gamma_{-n} = -\gamma_n$; $\alpha_{-n} = \bar{\alpha}_n$. If we could treat the right-hand side of (4) as if it were a finite sum, we could show, by Dirichlet's theorem on Diophantine approximation (or by the theory of almost-periodic functions), that any value taken by it must be approximately repeated for arbitrarily large values of u (of either sign); and we could deduce that any value taken by the function

$$f_0(x) = x^{-\frac{1}{2}}\{L_0(x) - I(x)\}$$

must be approximately repeated for arbitrarily large values of x (and again for arbitrarily small positive values of x). In fact, the Diophantine approximation (or almost-periodicity) can be pursued in both directions from any given value of u. Thus, if we could find a value of u for which

$$F_0(u) > 0, \tag{5}$$

we might expect to be able to deduce that

$$\limsup_{x\to\infty} x^{-\frac{1}{2}}\{L_0(x) - I(x)\} > 0, \qquad \limsup_{x\to 0+} x^{-\frac{1}{2}}\{L_0(x) - I(x)\} > 0;$$

or, since $x^{-\frac{1}{2}}I(x) \to 0$ as $x \to \infty$ and $L_0(x) = 0$ for $0 < x < 1$,

$$\limsup_{x\to\infty} x^{-\frac{1}{2}}L_0(x) > 0, \qquad \liminf_{x\to 0+} x^{-\frac{1}{2}}I(x) < 0.$$

Since $L_0(x)$ differs from $L(x)$ only when x is an integer, the difference then being $\pm\frac{1}{2}$, the former of these two inequalities could then be written

$$\limsup x^{-\frac{1}{2}}L(x) > 0 \quad (x \to \infty);$$

and this would contradict (1).

Since the right-hand side of (4) is in fact an infinite series, the purely formal argument is inapplicable as it stands since we cannot apply Dirichlet's theorem to an infinite number of γ_n (even if the series containing them is convergent, and we know little about this here). In this paper I use the Fejér kernel method, as set out in Ingham (**5**), to reduce the infinite series to a finite sum, and we are thus able to cast our arguments into a rigorous form. More precisely, I shall prove the following

theorem, in which α_n ($n = \pm 1, \pm 2, \pm 3, \ldots; \alpha_{-n} = \bar{\alpha}_n$) is the set of numbers occurring in the explicit formula (4).

THEOREM A. (i) *If the set of numbers α_n is unbounded, then*

$$\liminf x^{-\frac{1}{4}} L(x) = -\infty, \qquad \limsup x^{-\frac{1}{4}} L(x) = +\infty \quad (x \to \infty);$$

$$\liminf x^{-\frac{1}{4}} I(x) = -\infty, \qquad \limsup x^{-\frac{1}{4}} I(x) = +\infty \quad (x \to 0+).$$

(ii) *If the set of numbers α_n is bounded, then*

$$\liminf x^{-\frac{1}{4}} L(x) \leqslant x_0^{-\frac{1}{4}} \{L_0(x_0) - I(x_0)\} \leqslant \limsup x^{-\frac{1}{4}} L(x) \quad (x \to \infty),$$

$$\liminf x^{-\frac{1}{4}} I(x) \leqslant x_0^{-\frac{1}{4}} \{I(x_0) - L_0(x_0)\} \leqslant \limsup x^{-\frac{1}{4}} I(x) \quad (x \to 0+),$$

where x_0 is any fixed positive number.

From this we deduce

THEOREM B.
$$\liminf_{x \to \infty} x^{-\frac{1}{4}} L(x) = -\limsup_{x \to 0+} x^{-\frac{1}{4}} I(x),$$

$$\limsup_{x \to \infty} x^{-\frac{1}{4}} L(x) = -\liminf_{x \to 0+} x^{-\frac{1}{4}} I(x).$$

The object of these theorems is to enable us to apply to $L(x)$ any information that we may obtain about $I(x)$ (e.g. from its representation as an infinite series).

2. The proof of Theorem A (ii) will be based on (4) together with a property of the partial sums of the series on the right-hand side (Theorem 1). In establishing this property we can conveniently include a proof of (3)—or (4)—not given in detail in (1) owing to its similarity to the corresponding treatment of $M_0(x)$ by Titchmarsh (8). When we come to the proof of Theorem B, we shall find that we are able to deduce this theorem (without repeating the Diophantine-approximation argument) by first choosing a suitable x_0 and then applying Theorem A.

THEOREM 1. *Let X be any (fixed) positive number. Then, for $0 < x \leqslant X$,*

$$\left| \sum_{|\gamma_n| \leqslant T} \alpha_n x^{\rho_n} \right| \leqslant K,$$

where the numbers α_n and $\rho_n = \frac{1}{2} + \gamma_n i$ ($n = \pm 1, \pm 2, \ldots$) are those occurring in (4), $K = K(X)$ is a constant, and $T = T_m$ tends to infinity through a particular sequence.

Consider the integral

$$J = \frac{1}{2\pi i} \int_C Z(s) x^s \, ds, \qquad Z(s) = \frac{\zeta(2s)}{s\zeta(s)},$$

taken in the positive sense round the rectangle C with vertices $a - Ti$,

$2-Ti$, $2+Ti$, $a+Ti$, where a is fixed ($0 < a < \frac{1}{2}$) and $T \to \infty$ through a sequence T_m ($m = 1, 2, 3, \ldots$) such that

$$\frac{1}{\zeta(s)} = O(T^{\epsilon}) \quad (\tfrac{1}{2} \leqslant \sigma \leqslant 2;\ |t| = T = T_m),$$

where ϵ is any fixed positive number. This holds also for $-1 \leqslant \sigma \leqslant \frac{1}{2}$, $|t| = T = T_m$, by the functional equation. The existence of such a sequence T_m is proved (on the Riemann hypothesis) by Titchmarsh [(**8**) § 2 (2.2), and p. 20; also (**9**) 304 (14.16.2)].

By Cauchy's theorem of residues

$$J = \alpha_0 x^{\frac{1}{2}} + \sum_{|\gamma_n| \leqslant T} \alpha_n x^{\rho_n},$$

where $\alpha_0 = 1/\zeta(\frac{1}{2})$, as in (4).

Let C_1, C_2, C_3, C_4 be the four sides of C numbered in the positive sense from the lower horizontal side, and let J_1, J_2, J_3, J_4 be the contributions to J of these sides respectively. On C_1 and C_3

$$\frac{1}{\zeta(s)} = O(T^{\epsilon}), \qquad \zeta(2s) = O(T^{\frac{1}{2}+\epsilon}) \quad (a \leqslant \sigma \leqslant 2,\ |t| = T = T_m).$$

Also, taking $X \geqslant 1$ (as we may),

$$|x^s| = x^{\sigma} \leqslant X^{\sigma} \leqslant X^2 \quad (a \leqslant \sigma \leqslant 2;\ 0 < x \leqslant X).$$

Hence, on C_1 and C_3,

$$Z(s)x^s = \frac{1}{s}\zeta(2s)\frac{1}{\zeta(s)}x^s = O(T^{-1}T^{\frac{1}{2}+\epsilon}T^{\epsilon}X^2) = O(T^{-\frac{1}{2}+2\epsilon}X^2) = O(X^2), \tag{6}$$

when $T \to \infty$ through the sequence T_m. Since C_1 and C_3 are of fixed length $2-a$, it follows that

$$|J_1| < A_1 X^2, \qquad |J_3| < A_1 X^2.$$

For s on C_2, $$Z(s)x^s = \sum_{n=1}^{\infty} \lambda(n)\frac{(x/n)^s}{s},$$

and this series is uniformly convergent for s on C_2 (x fixed). Hence we may integrate term by term and obtain

$$J_2 = \sum_{n=1}^{\infty} \lambda(n)\frac{1}{2\pi i}\int_{C_2} \frac{(x/n)^s}{s}\,ds$$

$$= L_0(x) - \sum_{n=1}^{\infty} \lambda(n)\Delta\left(\frac{x}{n}, T\right) = L_0(x) - S, \text{ say},$$

where the notation is the same as in Ingham [(6) 75, Theorem G]. Now, for $x > 0$ [(6) 75, Theorem G, (10)],

$$|S| \leqslant \sum_{n=1}^{\infty} \left(\frac{x}{n}\right)^2 = A_2 x^2.$$

Hence, if $0 < x \leqslant X$,

$$|J_2| \leqslant |L_0(x)| + |S| \leqslant x + A_2 x^2 \leqslant X + A_2 X^2.$$

Finally [see (1) § 3, 96]

$$|J_4| = \left| \frac{1}{2\pi i} \int_{C_4} Z(s) x^s \, ds \right| \leqslant x^a \frac{1}{2\pi} \int_{C_4} |Z(s)\, ds|$$

$$\leqslant x^a \frac{1}{2\pi} \int_{(a)} |Z(s)\, ds| \leqslant A_3 X^a,$$

since the infinite integral is convergent.

Collecting these results, we obtain

$$|J| \leqslant |J_1| + |J_2| + |J_3| + |J_4|$$
$$< 2A_1 X^2 + X + A_2 X^2 + A_3 X^a.$$

Hence, comparing with the result given by Cauchy's theorem, we finally deduce

$$\left| \sum_{|\gamma_n| \leqslant T} \alpha_n x^{\rho_n} \right| \leqslant |\alpha_0| x^{\frac{1}{2}} + |J| < |\alpha_0| X^{\frac{1}{2}} + A_4 X^2 + X + A_3 X^a.$$

This proves Theorem 1.

To deduce (3), or (4), from our formulae, take a fixed $x > 0$ and let $T = T_m \to \infty$ in the identity

$$J_1 + J_2 + J_3 + J_4 = J.$$

First, $J_1 \to 0$ and $J_3 \to 0$ by (6). Next, $J_2 \to L_0(x)$, since the series S is uniformly convergent (with respect to T) for $T > 0$ and each term tends to 0 as $T \to \infty$ [(6) 75, Theorem G, (10) and (9)]. Finally $J_4 \to -I(x)$ by the definition of $I(x)$. We deduce (3), or (4), with the right-hand side of (4), for example, interpreted as the limit of

$$\alpha_0 + \sum_{0 < \gamma_n \leqslant T} (\alpha_n e^{\gamma_n u i} + \bar{\alpha}_n e^{-\gamma_n u i})$$

when $T \to \infty$ through the sequence T_m.

3. Proof of Theorem A (i)

The proof of Theorem A is divided into two parts corresponding to parts (i) and (ii) of the theorem. In part (i) of the theorem we use a classical argument due to E. Schmidt [see, e.g., Ingham (6) 91–2,

Theorem 33]; in part (ii) we adopt the standard method of Ingham (**5**), but with the k and k^* (or $\mathfrak{k}$ and $\mathfrak{K}$) used the other way round.

Let C be any real number. Then

$$\int_1^\infty \frac{L(x)+Cx^{\frac{1}{2}}}{x^{s+1}}\,dx = \frac{\zeta(2s)}{s\zeta(s)}+\frac{C}{s-\frac{1}{2}} \quad (\sigma > 1).$$

We write this as

$$\int_1^\infty \frac{L_C(x)}{x^s}\,dx = Z_C(s) \quad (\sigma > 1). \tag{7}$$

Now suppose that $L_C(x) \geqslant 0$ for, say, $x \geqslant X_0 > 1$. Then, by Landau's theorem, the line of convergence of the integral on the left-hand side of (7) is $\sigma = \frac{1}{2}$. For the singularities of $Z_C(s)$ are at $s = \frac{1}{2}$ (unless $C = -\alpha_0$), $s = 0$, and all the *complex* zeros of $\zeta(s)$, and the first of these we meet when we continue $Z_C(s)$ to the left along the real axis is the point $s = \frac{1}{2}$ (or $s = 0$). It follows that equation (7) is true for $\sigma > \frac{1}{2}$. Hence, for $\sigma > \frac{1}{2}$ and any real t,

$$|Z_C(\sigma+ti)| = \left|\int_1^\infty \frac{L_C(x)}{x^{\sigma+ti}}\,dx\right| \leqslant \int_1^\infty \frac{|L_C(x)|}{x^\sigma}\,dx = \int_1^{X_0} \frac{|L_C(x)|}{x^\sigma}\,dx + \int_{X_0}^\infty \frac{L_C(x)}{x^\sigma}\,dx$$

$$= \int_1^{X_0} \frac{|L_C(x)|-L_C(x)}{x^\sigma}\,dx + Z_C(\sigma) \leqslant \int_1^{X_0} 2|L_C(x)|\,dx + Z_C(\sigma) = K+Z_C(\sigma),$$

where $K = K(C, X_0) = K(C)$ is a constant. We thus have

$$|Z_C(\sigma+ti)| \leqslant K+Z_C(\sigma) \quad (\sigma > \tfrac{1}{2}). \tag{8}$$

Select a zero $\rho = \rho_n = \frac{1}{2}+\gamma_n i$ of $\zeta(s)$. In (8) take $t = \gamma_n$ and multiply through by $\sigma-\frac{1}{2}$. This gives

$$(\sigma-\tfrac{1}{2})|Z_C(\sigma+\gamma i)| \leqslant (\sigma-\tfrac{1}{2})K+(\sigma-\tfrac{1}{2})Z_C(\sigma) \quad (\sigma > \tfrac{1}{2}).$$

Now let $\sigma \to \frac{1}{2}+$. We then have

$$|\alpha_n| \leqslant \alpha_0+C.$$

The number C is at our disposal. Suppose it chosen so that

$$C < |\alpha_n|-\alpha_0.$$

Then the assumption that $L_C(x) \geqslant 0$ for all x sufficiently large leads to a contradiction. Hence $L_C(x) < 0$ for some arbitrarily large x; that is,

$$x^{-\frac{1}{2}}L(x) < -C$$

for some arbitrarily large x. Thus, when $x \to \infty$,

$$\liminf x^{-\frac{1}{4}} L(x) \leqslant -C.$$

Since C may be taken as near as we please to $|\alpha_n| - \alpha_0$, and since this number can, by hypothesis, be taken as large as we please, it follows that

$$\liminf x^{-\frac{1}{4}} L(x) = -\infty \quad (x \to \infty).$$

In the same way, by considering the function $-L(x) + Cx^{\frac{1}{4}}$ and taking $C < |\alpha_n| + \alpha_0$, we show that, when $x \to \infty$,

$$\limsup x^{-\frac{1}{4}} L(x) = +\infty \quad (x \to \infty).$$

This completes the proof of part (i) of Theorem A as far as the operation $x \to \infty$ is concerned.

For $x \to 0+$, we consider the integral

$$\int_0^\infty I\left(\frac{1}{x}\right) x^{s-1}\, dx$$

in place of

$$\int_1^\infty \frac{L(x)}{x^{s+1}}\, dx$$

and express the integral in terms of the zeta-function (e.g. by Mellin's inversion formula). We have, for $x > 0$,

$$I(x) = \frac{1}{2\pi i} \int_{(a)} \frac{\zeta(2s)}{s\zeta(s)} x^s\, ds = \frac{1}{2\pi i} \int_{(a)} \frac{\zeta(2s)}{s\zeta(s)} \left(\frac{1}{x}\right)^{-s} ds \quad (0 < a < \tfrac{1}{2}).$$

Also

$$\frac{1}{2\pi i} \int_{(a)} \frac{\zeta(2s)}{s\zeta(s)}\, ds \quad (0 < a < \tfrac{1}{2})$$

is absolutely convergent, and the integrand has a bounded derivative, and is therefore continuous and of bounded variation, in any finite interval of t. Hence, by the inversion formula of Mellin [see, e.g., Titchmarsh (**10**) 46, Theorem 29],

$$\frac{\zeta(2s)}{s\zeta(s)} = \int_0^\infty \frac{I(x)}{x^{s+1}}\, dx = \int_0^\infty I\left(\frac{1}{x}\right) x^{s-1}\, dx \quad (0 < \sigma < \tfrac{1}{2}),$$

these integrals being absolutely convergent since $I(x) = O(x^a)$ $(x > 0)$ for every fixed a in $0 < a < \frac{1}{2}$ [(**1**) § 3, 96]. Now let c be any real number. Then

$$\int_1^\infty \left\{ I\left(\frac{1}{x}\right) + cx^{-\frac{1}{4}} \right\} x^{s-1}\, dx = \frac{\zeta(2s)}{s\zeta(s)} - \int_0^1 I\left(\frac{1}{x}\right) x^{s-1}\, dx - \frac{c}{s - \frac{1}{4}} \quad (0 < \sigma < \tfrac{1}{2}).$$

The integral on the right is a bounded function of s in $0 < b \leqslant \sigma < \frac{1}{2}$. For, taking $0 < a < b$, we have $I(u) = O(u^a)$ for $u > 0$, so that

$$\int_0^1 I\left(\frac{1}{x}\right)x^{s-1}\,dx = O\left(\int_0^1 x^{-a+\sigma-1}\,dx\right) = O\left(\int_0^1 x^{-a+b-1}\,dx\right) = O(1)$$

for all s in $b \leqslant \sigma < \frac{1}{2}$, since $b-a > 0$.

We write this as

$$\int_1^\infty J_c(x)x^{s-1}\,dx = Y_c(s) \quad (0 < \sigma < \tfrac{1}{2}). \tag{9}$$

Now suppose that $J_c(x) \geqslant 0$ for, say, $x \geqslant x_0 > 1$. Then, exactly as before,

$$|Y_c(\sigma+ti)| \leqslant k+Y_c(\sigma) \quad (0 < \sigma < \tfrac{1}{2}), \tag{10}$$

where $k = k(c, x_0) = k(c)$ is a constant. Select a zero $\rho = \rho_n = \frac{1}{2}+\gamma_n i$ of $\zeta(s)$. In (10) take $t = \gamma_n$, multiply through by $\frac{1}{2}-\sigma$ (> 0), and let $\sigma \to \frac{1}{2}-$. We then have

$$|\alpha_n| \leqslant -\alpha_0+c.$$

The number c is at our disposal. Suppose it chosen so that

$$c < |\alpha_n|+\alpha_0.$$

Then the assumption that $J_c(x) \geqslant 0$ for all x sufficiently large leads to a contradiction. Hence $J_c(x) < 0$ for some arbitrarily large x. Since c may be taken as large as we please, we conclude that

$$\liminf x^{\frac{1}{4}}I(1/x) = -\infty \quad (x \to \infty);$$

or

$$\liminf x^{-\frac{1}{4}}I(x) = -\infty \quad (x \to 0+).$$

In the same way, by considering the function $-I(1/x)+cx^{-\frac{1}{4}}$ and taking $c < |\alpha_n|-\alpha_0$, we show that

$$\limsup x^{-\frac{1}{4}}I(x) = +\infty \quad (x \to 0+).$$

This completes the proof of part (i) of Theorem A.

4. Proof of Theorem A (ii)

Write

$$f_0(x) = x^{-\frac{1}{4}}\{L_0(x)-I(x)\} \quad (x > 0), \qquad F_0(u) = f_0(e^u).$$

Let $$k(x) = \begin{cases} 1-|x| & (|x| \leqslant 1), \\ 0 & (|x| > 1), \end{cases} \qquad k^*(y) = \left(\frac{\sin \frac{1}{2}y}{\frac{1}{2}y}\right)^2.$$

Then

$$\int_{-\infty}^{\infty} k(x)e^{\pm xyi}\,dx = k^*(y). \tag{11}$$

Now take a positive number η, and write

$$k_\eta(u) = \frac{1}{\eta} k\left(\frac{u}{\eta}\right).$$

Then $$k_\eta^*(y) = k^*(\eta y).$$

We start from the formula

$$\int_{\omega-\eta}^{\omega+\eta} k_\eta(u-\omega) F_0(u)\, du$$
$$= \alpha_0 \int_{\omega-\eta}^{\omega+\eta} k_\eta(u-\omega)\, du + \sum_{n=1}^{\infty} \int_{\omega-\eta}^{\omega+\eta} k_\eta(u-\omega)(\alpha_n e^{\gamma_n u i} + \bar{\alpha}_n e^{-\gamma_n u i})\, du.$$

This equation is obtained by multiplying the explicit formula (4) for $F_0(u)$ by $k_\eta(u-\omega)$ and integrating term by term from $u = \omega-\eta$ to $u = \omega+\eta$. The process is legitimate since $k_\eta(u-\omega)$ is integrable in the sense of Lebesgue and, by Theorem 1, a certain subsequence of the sequence of partial sums of the series occurring in the formula for $F_0(u)$ is boundedly convergent in the range $[\omega-\eta, \omega+\eta]$. Lebesgue's theorem then shows that the term-by-term integration is legitimate, if the resulting series is interpreted as the limit of the corresponding subsequence of partial sums. But then, in case (ii), this series is in fact convergent in the ordinary sense, since the nth term is $O(\gamma_n^{-2})$, as may be seen at once from the form (12) given below.

Now substitute $u = \omega+y$;

$$\int_{-\eta}^{\eta} k_\eta(y) F_0(\omega+y)\, dy$$
$$= \alpha_0 \int_{-\eta}^{\eta} k_\eta(y)\, dy + \sum_{n=1}^{\infty} \int_{-\eta}^{\eta} k_\eta(y)(\alpha_n e^{\gamma_n(\omega+y)i} + \bar{\alpha}_n e^{-\gamma_n(\omega+y)i})\, dy.$$

Since $k_\eta(y) = 0$ for $|y| > \eta$, we may replace the limits of integration on the right-hand side by $-\infty, \infty$. Hence

$$\int_{-\eta}^{\eta} k_\eta(y) F_0(\omega+y)\, dy = \alpha_0 k_\eta^*(0) + \sum_{n=1}^{\infty} (\alpha_n e^{\gamma_n \omega i} + \bar{\alpha}_n e^{-\gamma_n \omega i}) k_\eta^*(\gamma_n)$$
$$= \alpha_0 + \sum_{n=1}^{\infty} (\alpha_n e^{\gamma_n \omega i} + \bar{\alpha}_n e^{-\gamma_n \omega i}) \left(\frac{\sin \frac{1}{2}\eta\gamma_n}{\frac{1}{2}\eta\gamma_n}\right)^2. \quad (12)$$

We now replace the infinite series on the right-hand side of (12) by the corresponding finite sum extended over those γ_n for which $0 < \gamma_n \leqslant T$. Since $\alpha_n = O(1)$, by hypothesis, the error will be

$$O\left\{ \sum_{\gamma_n > T} |\alpha_n| \left(\frac{1}{\frac{1}{2}\eta\gamma_n}\right)^2 \right\} = O\left(\frac{1}{\eta^2} \sum_{\gamma_n > T} \frac{1}{\gamma_n^2}\right) = O\left(\frac{\log T}{T\eta^2}\right),$$

where the constants of the O's are absolute. [See, e.g., Ingham (6) 70, Theorem 25 b.] We thus have

$$\int_{-\eta}^{\eta} k_\eta(y)F_0(\omega+y)\,dy$$

$$= \alpha_0+\sum_{0<\gamma_n\leqslant T}(\alpha_n e^{\gamma_n\omega i}+\bar{\alpha}_n e^{-\gamma_n\omega i})\left(\frac{\sin\frac{1}{2}\eta\gamma_n}{\frac{1}{2}\eta\gamma_n}\right)^2+O\left(\frac{\log T}{T\eta^2}\right)$$

$$= S(\omega)+O\left(\frac{\log T}{T\eta^2}\right), \tag{13}$$

where

$$S(\omega) = S_{T,\eta}(\omega) = \alpha_0+\sum_{0<\gamma_n\leqslant T}(\alpha_n e^{\gamma_n\omega i}+\bar{\alpha}_n e^{-\gamma_n\omega i})\left(\frac{\sin\frac{1}{2}\eta\gamma_n}{\frac{1}{2}\eta\gamma_n}\right)^2.$$

Now take an integer $q \geqslant 3$, and choose, by Dirichlet's theorem on Diophantine approximation, a number $U = U(T, q)$ in the range

$$q \leqslant U \leqslant q^{N(T)+1} \tag{14}$$

so that

$$U\frac{\gamma_n}{2\pi} = l_n+\phi_n \quad \left(l_n \text{ integer; } |\phi_n| < \frac{1}{q};\ n = 1, 2,..., N(T)\right).$$

Then, for $n = 1, 2,..., N(T)$ and all real v,

$$|e^{\pm\gamma_n(U+v)i}-e^{\pm\gamma_n vi}| = |e^{\pm(\gamma_n v+2\pi\phi_n)i}-e^{\pm\gamma_n vi}| \leqslant |2\pi\phi_n| < \frac{2\pi}{q}.$$

Hence

$$|S(U+v)-S(v)| \leqslant \sum_{0<\gamma_n\leqslant T}\frac{4\pi}{q}|\alpha_n| = O\left(\frac{1}{q}\sum_{0<\gamma_n\leqslant T}1\right) = O\left(\frac{T\log T}{q}\right), \tag{15}$$

for all real v, since $\alpha_n = O(1)$ and $N(T) = O(T\log T)$. Hence, putting $\omega = U+u_0$ and $\omega = u_0$ in (13) and combining, we have

$$\int_{-\eta}^{\eta} k_\eta(y)\{F_0(U+u_0+y)-F_0(u_0+y)\}\,dy = S(U+u_0)-S(u_0)+O\left(\frac{\log T}{T\eta^2}\right)$$

$$= O\left(\frac{T\log T}{q}\right)+O\left(\frac{\log T}{T\eta^2}\right). \tag{16}$$

We now choose q, η to make this small when T is large, say

$$\eta = \eta_T = \frac{\log T}{T^{\frac{1}{2}}}, \qquad q = q_T = [T\log^2 T] \quad (T \text{ large}).$$

Then the right-hand side of (16) is $o(1)$ as $T \to \infty$, say less than $\frac{1}{2}\epsilon$ in modulus for all $T > T_0 = T_0(\epsilon)$ where ϵ is any fixed positive number. Since

$$\int_{-\eta}^{\eta} k_\eta(y)\,dy = k^*(0) = 1,$$

this gives

$$\left| \int_{-\eta}^{\eta} k_\eta(y)\{F_0(U+u_0+y)-F_0(u_0+y)\}\,dy \right| < \int_{-\eta}^{\eta} k_\eta(y)\tfrac{1}{2}\epsilon\,dy$$

$$(\eta = \eta_T;\ U = U(T,q) = U(T);\ T > T_0).$$

Since $k(x) \geqslant 0$ for all x, it follows that the expression inside $\{\}$ on the left-hand side is less than $\frac{1}{2}\epsilon$ for some y in $[-\eta, \eta]$ and also greater than $-\frac{1}{2}\epsilon$ for some y in $[-\eta, \eta]$:

$$F_0(U+u_0+y) < F_0(u_0+y)+\tfrac{1}{2}\epsilon$$

for some y in $[-\eta, \eta]$, and

$$F_0(U+u_0+y') > F_0(u_0+y')-\tfrac{1}{2}\epsilon$$

for some y' in $[-\eta, \eta]$. Substitute $y = u-u_0$, $y' = u'-u_0$;

$$F_0(U+u) < F_0(u)+\tfrac{1}{2}\epsilon \tag{17}$$

for some u in $[u_0-\eta, u_0+\eta]$, and

$$F_0(U+u') > F_0(u')-\tfrac{1}{2}\epsilon \tag{18}$$

for some u' in $[u_0-\eta, u_0+\eta]$.

Since $F_0(u)$ is a continuous function of u except at isolated points, we may suppose that $F_0(u)$ is continuous in the range $[u_0-\eta, u_0+\eta]$ except possibly at u_0. This is justified if η is small enough; that is, T large enough. Consider first the case in which $F_0(u)$ is continuous at u_0. Then we shall have

$$F_0(u_0)-\tfrac{1}{2}\epsilon < F_0(u) < F_0(u_0)+\tfrac{1}{2}\epsilon$$

for all u in $[u_0-\eta, u_0+\eta]$, if η is small enough. This combined with (17) and (18) gives

$$F_0(U+u) < F_0(u_0)+\epsilon$$

for some u in $[u_0-\eta, u_0+\eta]$, and

$$F_0(U+u') > F_0(u_0)-\epsilon$$

for some u' in $[u_0-\eta, u_0+\eta]$. Now choose $U = U_T$, say, to satisfy the Diophantine-approximation condition and so that, in addition, $U+u$ (or $U+u'$) is large and positive for all u (or u') in $[u_0-\eta, u_0+\eta]$. This choice of U is possible if $T > T_1 = T_1(\epsilon, u_0)$, since $U \geqslant q$. We then have

$$F_0(U+u) < F_0(u_0)+\epsilon$$

for some arbitrarily large $U+u$, and

$$F_0(U+u') > F_0(u_0)-\epsilon$$

for some arbitrarily large $U+u'$. Hence, keeping ϵ fixed and making $T \to \infty$, we deduce

$$\liminf F_0(u) \leqslant F_0(u_0)+\epsilon, \qquad \limsup F_0(u) \geqslant F_0(u_0)-\epsilon \quad (u \to \infty).$$

Since ϵ may be taken arbitrarily small, it follows that

$$\liminf F_0(u) \leqslant F_0(u_0), \qquad \limsup F_0(u) \geqslant F_0(u_0) \quad (u \to \infty);$$

that is,
$$\liminf F_0(u) \leqslant F_0(u_0) \leqslant \limsup F_0(u) \quad (u \to \infty). \tag{19}$$

To deduce the corresponding inequalities for the lower and upper limits as $u \to -\infty$, we now pursue the Diophantine approximation in the opposite (negative) direction. Thus, going back to (13), we now put $\omega = -U+u_0$ and $\omega = u_0$ and combine. Hence, arguing as before, we have, in place of (17) and (18),

$$F_0(-U+u) < F_0(u)+\tfrac{1}{2}\epsilon \tag{17}'$$

for some u in $[u_0-\eta, u_0+\eta]$, and

$$F_0(-U+u') > F_0(u')-\tfrac{1}{2}\epsilon \tag{18}'$$

for some u' in $[u_0-\eta, u_0+\eta]$. Assuming that $F_0(u)$ is continuous at u_0, we conclude, in the same way as before, that

$$\liminf F_0(u) \leqslant F_0(u_0) \leqslant \limsup F_0(u) \quad (u \to -\infty). \tag{20}$$

Now suppose that $F_0(u)$ is discontinuous at the given point u_0. Then

$$F_0(u_0-) < F_0(u_0) < F_0(u_0+) \quad \text{or} \quad F_0(u_0-) > F_0(u_0) > F_0(u_0+).$$

Hence, since $F_0(u)$ is continuous except at isolated points, we can find u_1 and u_2 sufficiently near to u_0 and on opposite sides of it such that $F_0(u)$ is continuous at u_1 and u_2 and also

$$F_0(u_1) \leqslant F_0(u_0) \leqslant F_0(u_2). \tag{21}$$

Now (19) and (20) are true for any u_0 at which $F_0(u)$ is continuous. They are therefore true with u_0 replaced by either u_1 or u_2. Applying (19), for instance, with u_0 replaced first by u_1 then by u_2, we have, when $u \to \infty$,

$$\liminf F_0(u) \leqslant F_0(u_1) \leqslant \limsup F_0(u),$$

$$\liminf F_0(u) \leqslant F_0(u_2) \leqslant \limsup F_0(u).$$

These, combined with (21), now give

$$\liminf F_0(u) \leqslant F_0(u_1) \leqslant F_0(u_0) \leqslant F_0(u_2) \leqslant \limsup F_0(u) \quad (u \to \infty).$$

We thus have

$$\liminf F_0(u) \leqslant F_0(u_0) \leqslant \limsup F_0(u) \quad (u \to \infty).$$

Similarly (20) gives

$$\liminf F_0(u) \leqslant F_0(u_0) \leqslant \limsup F_0(u) \quad (u \to -\infty).$$

This proves that (19) and (20) are true also in the case in which $F_0(u)$ is discontinuous at u_0. Hence they are true for any given u_0.

Since $$e^{-\frac{1}{2}u}I(e^u) \to 0 \quad \text{as } u \to \infty$$

and $$L_0(e^u) = 0 \quad (u < 0),$$

we deduce, putting $x = e^u$ and $x_0 = e^{u_0}$ in (19) and (20) and using the definition of $F_0(u)$, the desired inequalities as stated in part (ii) of the theorem. This completes the proof of Theorem A.

5. Proof of Theorem B

Let $F_0(u)$ be the function defined in § 4, and let

$$\limsup_{\substack{\inf \\ u \to -\infty}} F_0(u) = \begin{cases} \Lambda_- \\ \lambda_- \end{cases}, \qquad \limsup_{\substack{\inf \\ u \to +\infty}} F_0(u) = \begin{cases} \Lambda_+ \\ \lambda_+ \end{cases} \quad (\text{possibly } \pm\infty).$$

We have to prove that

$$\lambda_- = \lambda_+, \qquad \Lambda_- = \Lambda_+. \tag{22}$$

If the set of numbers α_n is unbounded, the results are certainly true by Theorem A (i). We may therefore suppose that the α_n are bounded. Now consider the left-hand equation of (22). If $\lambda_+ > \lambda_-$, we choose a number B so that $\lambda_+ > B > \lambda_-$. Then, since $B > \lambda_-$, we have

$$F_0(u) < B$$

for some arbitrarily large negative u. Select one such value of u and call it u_0. Then, by Theorem A (ii) in the form (19) (left-hand half), we have

$$\lambda_+ < B;$$

and this is impossible since B was chosen to satisfy $\lambda_+ > B$. If, on the contrary, $\lambda_+ < \lambda_-$, we have a similar contradiction by Theorem A (ii) in the form (20) (left-hand half). Hence the only conclusion is

$$\lambda_- = \lambda_+.$$

Similarly, using the right-hand halves of (19) and (20), we can prove the right-hand equation of (22). This completes the proof of Theorem B.

The author is greatly indebted to Mr. A. E. Ingham for the suggestion of the problem and also for much valuable discussion and help.

REFERENCES

1. A. Y. Fawaz, 'The explicit formula for $L_0(x)$', *Proc. London Math. Soc.* (3) 1 (1950), 86–103.
2. H. Gupta, 'On a table of values of $L(n)$', *Proc. Indian Acad. Sci.*, Sect. A. 12 (1940), 407–9.
3. —— 'A table of values of Liouville's function $L(t)$', *Res. Bull. East Panjab Univ.* (1950), 45–63.

4. A. E. Ingham, 'On two conjectures in the theory of numbers', *American J. of Math.* 64 (1942), 313–19.
5. —— 'A note on the distribution of primes', *Acta Arithmetica*, 1 (1935), 201–11.
6. —— *The Distribution of Prime Numbers* (Cambridge, 1932).
7. G. Pólya, 'Verschiedene Bemerkungen zur Zahlentheorie', *Jahresb. deutschen Math.-Ver.* 28 (1919), 31–40.
8. E. C. Titchmarsh, 'Some properties of the Riemann zeta-function', *Quart. J. of Math.* (Oxford), 14 (1943), 16–26.
9. —— *The Theory of the Riemann Zeta-function* (Oxford, 1951).
10. —— *Introduction to the Theory of Fourier Integrals* (Oxford, 1937).

A PROOF OF BLASCHKE'S THEOREM ON THE REULEAUX TRIANGLE

By H. G. EGGLESTON (*Swansea*)

[Received 27 February 1952]

It is well-known that of all plane convex sets of constant width d, the Reuleaux triangle, bounded by three circular arcs of radius d, has the least area. There are four proofs of this result due respectively to Lebesgue [(**3**) 132], Blaschke (**2**), Fujiwara (**4**), and A. E. Mayer (**5**). In this note a new proof is given which, like Lebesgue's depends upon the fact that about any curve of constant width d may be circumscribed a regular hexagon of width d [(**3**) 131].

Let Γ be a convex plane set of constant width d circumscribed by a regular hexagon $ABCDEF$ of width d, and suppose that the notation is such that the order of $ABCDEFA$ is in the counter-clockwise sense of rotation about the hexagon. Take O, the centre of the hexagon, as the origin and the line through O parallel to AB as the initial line. Let $H(\theta)$ be the support function of Γ and $\rho(\theta)$ the radius of curvature of Γ at the point where the support line makes an angle θ with the initial line. In measuring this angle the support line is supposed to be directed in the counter-clockwise sense of rotation about Γ; $\rho(\theta)$ is zero at a vertex of Γ. Similarly define $H_1(\theta)$, $\rho_1(\theta)$ with respect to the Reuleaux triangle Γ_1 which can be inscribed in $ABCDEF$ to touch it at B, D, and F.

If two points of Γ lie on a circle C' of radius d, then the whole of the smaller arc of C' joining these two points also belongs to Γ.† If we apply this fact to the arcs BD, DF, FB, bounding Γ_1 we see that

$$H(\theta) \geqslant H_1(\theta) \quad (\tfrac{1}{3}\pi \leqslant \theta \leqslant \tfrac{2}{3}\pi;\ \text{or}\ \pi \leqslant \theta \leqslant \tfrac{4}{3}\pi;\ \text{or}\ \tfrac{5}{3}\pi \leqslant \theta \leqslant 2\pi).$$

Similarly from a consideration of the Reuleaux triangle obtained by reflecting Γ_1 in O, it follows that

$$H(\theta+\pi) \geqslant H_1(\theta) \quad (\tfrac{1}{3}\pi \leqslant \theta \leqslant \tfrac{2}{3}\pi;\ \text{or}\ \pi \leqslant \theta \leqslant \tfrac{4}{3}\pi;\ \text{or}\ \tfrac{5}{3}\pi \leqslant \theta \leqslant 2\pi).$$

The area of Γ is

$$\begin{aligned}
&\tfrac{1}{2}\int_0^{2\pi} \rho(\theta)H(\theta)\,d\theta \\
&\qquad = \tfrac{1}{2}\left\{\int_{\frac{1}{3}\pi}^{\frac{2}{3}\pi} + \int_{\pi}^{\frac{4}{3}\pi} + \int_{\frac{5}{3}\pi}^{2\pi}\right\}(\rho(\theta)H(\theta)+\rho(\theta+\pi)H(\theta+\pi))\,d\theta \\
&\qquad \geqslant \tfrac{1}{2}\left\{\int_{\frac{1}{3}\pi}^{\frac{2}{3}\pi} + \int_{\pi}^{\frac{4}{3}\pi} + \int_{\frac{5}{3}\pi}^{2\pi}\right\}(\rho(\theta)+\rho(\theta+\pi))H_1(\theta)\,d\theta.
\end{aligned}$$

† See (**1**) 82; Lemma 1, Corollary 2 or 5.

But $\rho(\theta)+\rho(\theta+\pi) = d$.† Thus this last expression is equal to

$$\tfrac{1}{2}\left\{\int_{\frac{1}{2}\pi}^{\frac{3}{2}\pi} + \int_{\pi}^{\frac{3}{2}\pi} + \int_{\frac{3}{2}\pi}^{2\pi}\right\} d\,H_1(\theta)\,d\theta$$

$$= \tfrac{1}{2}\int_0^{2\pi} \rho_1(\theta)\,H_1(\theta)\,d\theta,$$

and this last integral is the area of Γ_1.

This proves the result.

† See (3) 128.

REFERENCES

1. A. S. Besicovitch, 'Measure of asymmetry of convex curves (II). Curves of constant width', *J. of London Math. Soc.* (2) 26 (1951), 81–93.
2. W. Blaschke, 'Konvexe Bereiche gegebener konstanter Breite und kleinsten Inhalts', *Math. Annalen* 76 (1915) 504–13.
3. T. Bonnesen and W. Fenchel, *Theorie der Konvexen Körper* (Berlin, 1934).
4. M. Fujiwara, 'Analytical proof of Blaschke's theorem on the curve of constant breadth with minimum area (I), (II)', *Proc. Imp. Acad. Japan* 3 (1927) 307–9; 7 (1931) 300–2.
5. A. E. Mayer, 'Der Inhalt der Gleichdicke', *Math. Annalen*, 110 (1934–5), 97–127.

ON INFINITELY MULTICOHERENT SPACES

By A. H. STONE (*Manchester*)

[Received 28 February 1952]

1. Introduction

THE 'degree of multicoherence', $r(S)$, of a connected topological space S is defined by

$$r(S) = \sup b_0(A \cap B), \tag{1}$$

the supremum being taken over all pairs of *closed* connected sets A, B such that $A \cup B = S$. Here $b_0(X)$ denotes the number of components of X less one (or ∞ if this number is infinite). If now $r(S) = \infty$, it is natural to ask if the bound in (1) is attained, i.e. if closed connected sets A, B exist such that $A \cup B = S$ and $b_0(A \cap B) = \infty$. If this is the case (as it is with the most familiar examples), we shall say that '$r(S)$ is attained'. Spaces for which $r(S)$ is *not* attained present a certain interest since in them watered-down versions of many properties of finitely multicoherent spaces will persist.† I shall show below (Theorem 2) that $r(S)$ is always attained if S is a one-dimensional Peano space, and shall give an example (§ 6), after some preliminary theorems, of a two-dimensional Peano space S for which $r(S)$ is not attained.

2. Open sets

It is known [see (**3**), 403] that in (1) we may take the sets A and B to be *open* instead of closed (provided that S is normal and locally connected). The corresponding question of 'attainment' is easily settled, and we dispose of it first.

THEOREM 1. *If A and B are open connected subsets of a Peano space S such that $A \cup B = S$, then $b_0(A \cap B) < \infty$.*

Suppose the contrary. Since

$$\mathrm{Fr}(A) \cap \mathrm{Fr}(B) \subset (S-A) \cap (S-B) = 0,$$

there exist disjoint open sets U and V containing $\mathrm{Fr}(A)$ and $\mathrm{Fr}(B)$ respectively. Let $\{C_\lambda\}$ be the (infinite) collection of components of $A \cap B$. Each $\mathrm{Fr}(C_\lambda)$ meets both $\mathrm{Fr}(A)$ and $\mathrm{Fr}(B)$, and C_λ is connected; hence, for each λ, there exists a point $p_\lambda \in C_\lambda - (U \cup V)$. There exists a convergent infinite sequence (say) $p_{\lambda_n} \to q$ $(n = 1, 2, \ldots)$. Thus

$$q \in (\bar{A} \cap \bar{B}) - (U \cup V) \subset A \cap B,$$

† For example, if $r(S)$ is not attained, and A is any subset of S with $b_0(\bar{A}) < \infty$, then (if S is locally connected and normal) all but at most a finite number of the components of $S-A$ will have connected frontiers; cf. (**5**).

and therefore $q \in C_\lambda$ for some λ. The open set C_λ must then meet infinitely many distinct components C_{λ_n}, which is impossible.†

COROLLARY. *If G_1, G_2 are open subsets of a Peano space such that their frontiers are disjoint‡ and $b_0(G_1)+b_0(G_2) < \infty$, then $b_0(G_1 \cap G_2) < \infty$.*

For it clearly suffices to prove this assuming G_1 and G_2 to be connected. But [(3) 401] there then exist connected open sets $A \supset G_1$, $B \supset G_2$, such that $A \cup B = S$, and $G_1 \cap G_2$ is both open and closed relative to $A \cap B$; and the result follows.

3. One-dimensional spaces

THEOREM 2. *If S is a one-dimensional Peano space, and $r(S) = \infty$, then $r(S)$ is attained.*

The proof requires the following two lemmas; for the first, see (2), p. 214. The proof of the second is easy, and is omitted.

LEMMA 1. *If T is a regular curve, and $\epsilon > 0$ is given, there exist regular curves $E_1, E_2, \ldots, E_n$ such that $\bigcup E_i = T$, $\delta(E_i) < \epsilon$, $\mathrm{Fr}(E_i)$ (relative to T) is a finite set, the sets $\mathrm{Int}(E_i)$ are pairwise disjoint, and no three distinct sets E_i, E_j, E_k have a common point.*

LEMMA 2. *If E is any subset of a connected topological space T, and K is a connected subset of $\bar{E}$ which contains $\mathrm{Fr}(E)$, then $(T-E) \cup K$ is connected.*

Now suppose that S is a one-dimensional Peano space for which Theorem 2 is false. I first show that S is a regular curve. To do this, consider any true cyclic element C of S. If M and N are any continua such that $M \cup N = C$, we have

$$b_0(M \cap N) < \infty; \tag{2}$$

for there exists a monotone retraction (say) f of S onto C, and, if (2) were false, the sets $f^{-1}(M)$ and $f^{-1}(N)$ would be continua covering S and intersecting in infinitely many components, contrary to hypothesis. Now let p be any point of C, and let U be any neighbourhood of p. Since p is not a cut-point of the Peano space C, we readily construct a continuum $F \subset C$ such that $p \in C-F \subset U$. [See for example (6) 50.] Since further C is one-dimensional, we can cover F by a finite number

† This argument clearly applies to any normal, countably compact, connected and locally connected topological space S. It is easy to see from examples that these conditions cannot be greatly relaxed.

‡ The condition $\mathrm{Fr}(G_1) \cap \mathrm{Fr}(G_2) = 0$ can be weakened to

$$\mathrm{Fr}(G_1) \cap \mathrm{Fr}(G_2) \cap \mathrm{Fr}(G_1 \cap G_2) = 0;$$

this can be proved by refining the argument in (3) Theorem 8.

of connected open sets $W_1,\ldots, W_n$ (relative to C), each meeting F and such that $p \in C-\overline{W}_i$ and $\overline{W}_i \cap (C-W_i)$ is zero-dimensional. Let $W = \bigcup W_i$ and take V to be the component of $C-\overline{W}$ which contains p; then $p \in V \subset U$, V is open relative to C, V and $C-V$ are connected, and $\overline{V}-V$ is zero-dimensional. But, from (2) applied to $\overline{V}$ and $C-V$, $\overline{V}-V$ must now be finite; and this shows that C is a regular curve. This holds for every true cyclic element of S, and therefore [(6) 83] S is itself a regular curve.

Since $r(S) = \infty$, there exist continua A_1, B_1 such that $A_1 \cup B_1 = S$ and $b_0(A_1 \cap B_1) \geqslant 1$. Choose ϵ_1 positive and small enough, and apply Lemma 1 to cover S by regular curves $E_{11}, E_{12},\ldots, E_{1n_1}$, with diameters less than ϵ_1 and with the other properties specified in the lemma. Write C_1 for the union of the sets E_{1i} which meet A_1, and similarly D_1 for the union of the sets E_{1i} which meet B_1; clearly C_1, D_1 are continua covering S and such that $b_0(C_1 \cap D_1) < \infty$. Further, if ϵ_1 was chosen small enough, we have $C_1 \cap D_1 = H_{11} \cup H_{12}$, where H_{11}, H_{12} are closed and non-empty and $\rho(H_{11}, H_{12}) > \epsilon_1$. (In fact, H_{11}, H_{12} will be in preassigned neighbourhoods of sets similarly splitting $A_1 \cap B_1$.) Now, from Theorem 7 of (4) at least one of the sets E_{1i}, say E_{11}, satisfies $r(E_{11}) = \infty$. If E_{11} is not in $C_1 \cap D_1$, we 'join' it to $C_1 \cap D_1$ by a 'chain' of sets E_{1i}: that is, we find a sequence of distinct sets E_{11}, E_{1j_1}, $E_{1j_2},\ldots, E_{1j_m}$, of which every two consecutive members meet, and the last of which meets $C_1 \cap D_1$. Take such a chain with as few members as possible, and let J_1 denote the union of its members. Then J_1 meets only one of H_{11}, H_{12} (for only the last 'link' E_{1j_m} can meet either, and it is too small to meet both); and, on writing $X_1 = C_1 \cup J_1$, $Y_1 = D_1 \cup J_1$, we have that X_1, Y_1 are continua covering S, and $1 \leqslant b_0(X_1 \cap Y_1) < \infty$. Further, $X_1 \cap Y_1$ contains the regular curve E_{11}, which we rename S_1; thus $r(S_1) = \infty$, and $\mathrm{Fr}(S_1)$ is finite, say consisting of the k_1 points p_{11}, $p_{12},\ldots, p_{1k_1}$.

The foregoing is the first stage in an inductive construction; the general process will be sufficiently illustrated by the next stage. Since $r(S_1) = \infty$, there exist continua A_2, B_2 such that $A_2 \cup B_2 = S_1$ and $b_0(A_2 \cap B_2) \geqslant 2k_1$. Choose $\epsilon_2 > 0$ small enough, and apply Lemma 1 to cover S_1 by regular curves E_{21}, $E_{22},\ldots, E_{2n_2}$ with $\delta(E_{2i}) < \epsilon_2$. As before, let C_2, D_2 be the unions of the sets E_{2i} meeting A_2, B_2 respectively; then C_2, D_2 are continua such that $C_2 \cup D_2 = S_1$, and $C_2 \cap D_2$ is a union of $2k_1+1$ closed non-empty sets H_{2j} $(1 \leqslant j \leqslant 2k_1+1)$ every two of which are at a distance exceeding ϵ_2 from each other. At least one of the sets E_{2i}, say E_{21}, has $r(E_{21}) = \infty$; and this we join to $C_2 \cap D_2$ by a connected chain J_2 of sets E_{2i} such that J_2 meets only one set H_{2j}. Similarly each of the points p_{1k}

$(1 \leqslant k \leqslant k_1)$ is joined to $C_2 \cap D_2$ by a chain L_{2k} of sets E_{2i} meeting only one H_{2j}. Write

$$X_2 = C_2 \cup J_2 \cup \bigcup L_{2k}, \qquad Y_2 = D_2 \cup J_2 \cup \bigcup L_{2k};$$

then X_2, Y_2 are again continua such that $X_2 \cup Y_2 = S_1$, and $X_2 \cap Y_2$ contains both E_{21} and $\mathrm{Fr}(S_1)$. Moreover, $X_2 \cap Y_2 = (C_2 \cap D_2) \cup J_2 \cup \bigcup L_{2k}$; and, since $J_2 \cup \bigcup L_{2k}$ can meet at most k_1+1 of the sets H_{2j}, it follows that $k_1 \leqslant b_0(X_2 \cap Y_2) < \infty$. The set E_{21} is renamed S_2; its frontier (say) F_2 relative to S_1 is finite, and, since $\mathrm{Fr}(S_2) \subset F_2 \cup \mathrm{Fr}(S_1)$, it follows that $\mathrm{Fr}(S_2)$ (relative to S) is also finite, with (say) k_2 points $p_{21},..., p_{2k_2}$; and so on. In this way we obtain a sequence of regular curves S_n, with

$$S = S_0 \supset S_1 \supset ..., \quad \text{and} \quad r(S_n) = \infty,$$

and sequences of continua X_n, Y_n such that $X_n \cup Y_n = S_{n-1}$,

$$X_n \cap Y_n \supset S_n \cup \mathrm{Fr}(S_{n-1}), \quad \text{and} \quad k_{n-1} \leqslant b_0(X_n \cap Y_n) < \infty,$$

where k_n is the number of points in $\mathrm{Fr}(S_n)$ $(n = 1, 2,...)$.

Now define recursively $P_1 = (X_1 - S_1) \cup X_2$, $Q_1 = (Y_1 - S_1) \cup Y_2$, $P_n = (P_{n-1} - S_n) \cup X_{n+1}$, $Q_n = (Q_{n-1} - S_n) \cup Y_{n+1}$. It is easy to show by induction that P_n and Q_n are closed, $P_n \cup Q_n = S$, $P_n \cap Q_n \supset S_{n+1}$, and (using Lemma 2) that P_n and Q_n are connected. Further, we have

$$P_n \cap Q_n = \{P_{n-1} \cap Q_{n-1} - \mathrm{Int}(S_n)\} \cup (X_{n+1} \cap Y_{n+1}).$$

Now, one of the components of $P_{n-1} \cap Q_{n-1}$ contains S_n; the other components of $P_{n-1} \cap Q_{n-1}$ are components of $P_{n-1} \cap Q_{n-1} - \mathrm{Int}(S_n)$ which cannot meet $X_{n+1} \cap Y_{n+1}$; and $X_{n+1} \cap Y_{n+1}$ has at least k_n+1 components, at most k_n of which can meet $P_{n-1} \cap Q_{n-1} - \mathrm{Int}(S_n)$. Hence

$$b_0(P_n \cap Q_n) \geqslant b_0(P_{n-1} \cap Q_{n-1})+1,$$

from which we readily obtain $b_0(P_n \cap Q_n) \geqslant n$.

Clearly $P_{n-1} \supset P_n$ (for $X_{n+1} \subset S_n \subset X_n \subset P_{n-1}$), and similarly $Q_{n-1} \supset Q_n$; hence, defining $A = \bigcap P_n$, $B = \bigcap Q_n$, we have that A, B are continua and $A \cup B = S$. Further, it is easy to see (by induction over m) that, if $m \geqslant n$, then $P_m \cap Q_m \supset (P_n \cap Q_n) - \mathrm{Int}(S_{n+1})$; thus $P_m \cap Q_m$, and so also $A \cap B$, meets each component of $P_n \cap Q_n$. Thus $b_0(A \cap B) \geqslant n$ for every n, so that $b_0(A \cap B) = \infty$; and this contradiction establishes the theorem.

4. Infinite-dimensional spaces

In this section I shall obtain a class of infinite-dimensional spaces S for which $r(S)$ is infinite but not attained; a two-dimensional example will be derived later (§ 6). First I recapitulate briefly some results, mostly due to Eilenberg (**1**), on which the proofs will be based.

Let S be any continuum (for simplicity), and consider the family S^{1S} of (continuous) mappings f of S in the circle S^1 of complex numbers of unit modulus. If $f, g \in S^{1S}$, and, if for some $X \subset S$ there exists a (continuous) real-valued function ϕ defined on X such that

$$f(x) = \exp(i\phi(x))g(x) \quad \text{on } X,$$

we write $f \sim g$ on X; in particular, if $f(x) = \exp(i\phi(x))$ on X, we write $f \sim 1$ on X. Defining $f.g$ by $(f.g)(x) = f(x)g(x)$, where $f, g \in S^{1S}$ (so that $f(x)$ and $g(x)$ are complex numbers), we see that S^{1S} is an abelian group,† and that the mappings which ~ 1 on S form a subgroup, denoted by $P(S)$. The factor group $S^{1S}/P(S)$ is denoted by $B_1(S)$; it is an abelian group without torsion. If X, Y are closed subsets of S such that $X \cup Y = S$, the mappings $f\colon S \to S^1$ which ~ 1 both on X and on Y form a subgroup $P(X, Y)$ of S^{1S}; the factor-group $P(X, Y)/P(S)$ is a subgroup of $B_1(S)$, and its rank is denoted by $p(X, Y)$. The least upper bound of $p(X, Y)$, for all pairs of closed sets covering S, is denoted by $\rho(S)$; if S is a Peano space, then $\rho(S) = r(S)$. Finally, if X, Y are continua such that $X \cup Y = S$, then $p(X, Y) \geqslant b_0(X \cap Y)$. Thus, if S is a Peano space such that $r(S)$ is infinite and is attained, then certainly $\rho(S)$ is also 'attained', i.e. X, Y exist with $p(X, Y) = \infty$. Hence the following theorem settles the question of this paper so far as infinite-dimensional spaces are concerned.

THEOREM 3. *If S is the Cartesian product of a sequence of (non-empty) continua S_n, then $\rho(S) = \sup \rho(S_n)$; and, if each $\rho(S_n)$ is finite and $\rho(S) = \infty$, then $\rho(S)$ is not attained.*‡

We require the following lemma.

LEMMA 3. *If $S = \prod_{n=1}^{\infty} S_n$, then $B_1(S)$ is isomorphic to the restricted direct product $\prod_{n=1}^{\infty} B_1(S_n)$.*

To set up the isomorphism, consider first a mapping $f_n\colon S_n \to S^1$, and define $f_n^*\colon S \to S^1$ by $f_n^*(x_1, x_2, \ldots) = f_n(x_n)$; it follows trivially that, if $f_n \sim g_n$ on S_n, then $f_n^* \sim g_n^*$ on S, so the class $\{f_n\} \in B_1(S_n)$ is assigned

† Though all groups which arise here will be abelian, it is convenient to write them multiplicatively.

‡ This theorem and its proof are fairly immediate extensions of results of Eilenberg's [(**1**), 166, 168]. I give the argument, since it will be needed in § 6. There would be no difficulty in extending Theorem 3 to products of arbitrary many continua.

a unique class $\{f_n^*\} \in B_1(S)$. To the element $F = \{\{f_1\}, \{f_2\}, \ldots\}$ of $\prod B_1(S_n)$, where $f_n \sim 1$ (on S_n) for all $n > N(F)$, we assign the class

$$\theta F = \{f_1^* f_2^* \ldots f_N^*\} \in B_1(S);$$

clearly θ is a homomorphism of $B_1(S_n)$ in $B_1(S)$.

To define the inverse mapping, pick a fixed point $\xi = (\xi_1, \xi_2, \ldots)$ in S, and, for each f: $S \to S^1$, define f_n: $S_n \to S^1$ by

$$f_n(x_n) = f(\xi_1, \xi_2, \ldots, \xi_{n-1}, x_n, \xi_{n+1}, \ldots);$$

from the uniform continuity of f it follows that, if $n > N(f)$, the set $f_n(S_n)$ is small, so that $f_n \sim 1$ on S_n if $n > N$ [(1) 156]. If $f \sim g$ on S, it follows at once that $f_n \sim g_n$ on S_n, and on setting

$$\varpi\{f\} = \{\{f_1\}, \{f_2\}, \ldots, \{f_N\}\} \in \prod B_1(S_n),$$

we see that ϖ is a homomorphism of $B_1(S)$ in $\prod B_1(S_n)$.

It is a routine matter to verify that $\varpi\theta$ is the identity mapping. To show that $\theta\varpi$ is also an identity mapping comes to showing that, if $f \in S^{1S}$ and f^n: $S \to S^1$ is defined by

$$f^n(x_1, x_2, \ldots) = f(\xi_1, \ldots, \xi_{n-1}, x_n, \xi_{n+1} \ldots) \quad (\text{i.e. } f^n = f_n^*),$$

so that $f^n \sim 1$ for $n > N$, then $f \sim f^1 f^2 \ldots f^N$ on S. Consider f^{12}: $S \to S^1$ where $f^{12}(x_1, x_2, \ldots) = f(x_1, x_2, \xi_3, \ldots, \xi_n, \ldots)$; we have $f^{12} \sim f^1 f^2$ on the section $S_1 \times (\xi_2, \xi_3, \ldots)$ and also on the section $\xi_1 \times S_2 \times (\xi_3, \ldots)$, and therefore [(1) 157] $f^{12} \sim f^1 f^2$ on S. This argument can be repeated, and gives by induction that $f^1 f^2 \ldots f^N \sim f^{12\ldots N}$ on S, where

$$f^{12\ldots N}(x_1, x_2, \ldots) = f(x_1, \ldots, x_N, \xi_{N+1}, \ldots).$$

Here N can be arbitrarily large; if it is large enough, the uniform continuity of f gives $f \sim f^{12\ldots N}$ on S, whence $\theta\varpi\{f\} = \{f\}$. Thus θ and ϖ are isomorphisms, as required.

5. Proof of Theorem 3

Since S can be projected onto each S_n, we trivially have $\rho(S) \geqslant \rho(S_n)$ and so $\rho(S) \geqslant \sup \rho(S_n)$. Now consider the case in which $\sup \rho(S_n) = \infty$, but each $\rho(S_n)$ is finite; we must show that $\rho(S)$ is not attained. Suppose, on the contrary, there exist closed sets A, B such that $A \cup B = S$ and the group $Q = P(A, B)/P(S)$ has arbitrarily many (linearly) independent elements. From Lemma 3, Q, as a subgroup of $B_1(S)$, has an isomorphic image ϖQ in the restricted product $\prod B_1(S_n)$. Choose an element $F^0 = \{\{f_1^0\}, \{f_2^0\}, \ldots, \{f_N^0\}\}$ of ϖQ different from the unit element. I shall

show that ϖQ must also contain an element $G^0 \neq 1$ of the form $\{1, 1, \ldots, 1, \{g^0_{N+1}\}, \ldots\}$, i.e. with $\{g^0_n\} = 1$ for all $n \leqslant N$.

For otherwise the homomorphism τ: $\varpi Q \to \prod_1^N B_1(S_n)$ defined by $\tau\{\{f_1\}, \{f_2\}, \ldots\} = \{\{f_1\}, \ldots, \{f_N\}\}$ would be an isomorphism into, so that Q would be isomorphic to a subgroup of $B_1(Z)$ where Z is the subset $S_1 \times S_2 \times \ldots \times S_N \times (\xi_{N+1}, \ldots)$ of S; and it is not hard to see that, in this isomorphism, each $\{f\} \in Q$ corresponds to the class $\{f \mid Z\}$ (of f restricted to Z). Thus, since $f \sim 1$ on A and on B, Q would be isomorphic to a subgroup of $P(A \cap Z, B \cap Z)/P(Z)$, giving $\rho(Z) = \infty$. But [(**1**) 168]

$$\rho(Z) = \max\{\rho(S_1), \rho(S_2), \ldots, \rho(S_n)\},$$

which is finite; and this contradiction shows that the desired element G^0 of ϖQ exists.

Now write $$X = \prod_1^N S_n, \qquad Y = \prod_{N+1}^\infty S_n;$$

then $X \times Y = S$, and the mapping class θF^0 contains a mapping F of $X \times Y$ in S^1 such that $F(x, y)$ is independent of $y \in Y$ and, considered as a mapping of X in S^1, is not ~ 1 on X. Similarly θG^0 contains a mapping G: $X \times Y \to S^1$ which is independent of $x \in X$ and is not ~ 1 on Y. From [(**1**) 167], F and G cannot be '2-compatible'† on $X \times Y$, which contradicts the fact that θF^0 and θG^0 are in Q; and this contradiction proves that $\rho(S)$ is not attained.

Finally, if $\sup \rho(S_n) < \infty$, an entirely similar argument proves that $\rho(S) \leqslant \sup \rho(S_n)$.

Remark. Essentially the same argument also shows that, in Theorem 3, $\rho(S)$ is not attained even if some of the factors S_n have $\rho(S_n) = \infty$, provided $\rho(S_n)$ is not attained for any of these factors.

6. A two-dimensional example

Theorem 3 can be used to derive a class of two-dimensional Peano spaces T for which $\rho(T)$ is infinite but not attained. For simplicity we consider only a special case. Let G_n be a linear graph formed from n simple closed curves having a single common point ξ_n, but otherwise disjoint. Then $\rho(G_n) = r(G_n) = n$. Let $S = \prod_1^\infty G_n$; then S is a Peano space and, from Theorem 3, $\rho(S)$ is infinite but not attained. Now let T be the subset of S consisting of all points $x = \{x_n\}$ such that $x_n = \xi_n$ for all but at most two values of n. It is easy to see that T is a Peano space, so that $\rho(T) = r(T)$; further, T is the union of a countable family

† Mappings f, g: $S \to S^1$ are '2-compatible' on S if $S = A \cup B$ where A, B are closed and both f and $g \sim 1$ on both A and B.

of tori, and so is two-dimensional. There is an obvious monotone mapping of T onto each G_n; hence $r(T) \geqslant n$ for each n, so that $r(T) = \infty$. We have only to prove that $\rho(T)$ is not attained.

The first step is to prove that $B_1(S)$ and $B_1(T)$ are isomorphic in a natural way. This comes to proving the following two properties.

(1) Each $f\colon T \to S^1$ can be extended to a mapping $f^*\colon S \to S^1$.

For the uniform continuity of f shows that, if N is large enough (depending on f), $f = \exp(i\phi)g$ where, for $x = (x_1, x_2, \ldots)$ in T,

$$g(x) = f(x_1, \ldots, x_N, \xi_{N+1}, \ldots)$$

and where ϕ is a continuous real-valued function on T. There exists a (real-valued) extension ϕ^* of ϕ to S. To see that g can also be extended to S, it suffices to consider the finite product $G_1 \times \ldots \times G_N$. Since each G_n can be regarded as formed by n segments with their ends all identified, the product $G_1 \times \ldots \times G_N$ can be regarded as a cell-complex in which the cells are products of segments (with identifications on their boundaries); and g is defined on precisely the set of two-dimensional cells of this complex. Since the homotopy groups of S^1 vanish for dimensions $\geqslant 2$, g may be extended first to the 3-cells, and then to the cells of successively higher dimensions; finally g is extended to $G_1 \times \ldots \times G_N$, and so to a mapping $g^*\colon S \to S^1$. We merely define $f^* = \exp(i\phi^*)g^*$ on S.

(2) Let f^*, g^* be mappings of S in S^1, and let

$$f = f^* \mid T, \qquad g = g^* \mid T.$$

If $f \sim g$ on T, then $f^* \sim g^*$ on S.

It suffices to show that, if $f^* = 1$ on T, then $f^* \sim 1$ on S. Here again we may replace S by the cell-complex $G_1 \times \ldots \times G_N$, N $(= N(f))$ being large enough, and have $f^* = 1$ on the two-dimensional cells. It is not hard to show by induction that $f^* \sim 1$ on the set of cells of dimensions not exceeding n $(n = 2, 3, \ldots, N)$.

Thus the correspondence between $\{f^*\} \in B_1(S)$ and $\{f\} \in B_1(T)$ given by $f = f^* \mid T$ is an isomorphism onto. It has the further property:

(3) *f^* and g^* are 2-compatible† on S if and only if f and g are 2-compatible on T.*

For, if f^* and g^* are 2-compatible on S, they are trivially 2-compatible when restricted to T. Conversely, suppose that f and g are 2-compatible on T. By a theorem of Eilenberg [(1) 164] the mapping $(f, g)\colon T \to S^1 \times S^1$ is 'inessential', i.e. $f \sim f'$ and $g \sim g'$ on T where $(f', g')(T)$ is a subset of $S^1 \times S^1$ which is homeomorphic to the graph G_2. Call this subset G_2';

† As defined in the footnote on p. 304.

since G_2' also has zero homotopy groups in dimensions not less than 2, the same argument as before shows that (f', g') has an extension (f'^*, g'^*) mapping S in $G_2' \subset S^1 \times S^1$. Thus [(1) 164] f'^*, g'^* are 2-compatible mappings of S in S^1 which extend f', g' respectively. From (2) above, $f'^* \sim f^*$ and $g'^* \sim g^*$ on S, so that f^* and g^* are also 2-compatible on S.

Now suppose that closed sets A, B exist such that $A \cup B = T$ and the group $Q = P(A, B)/P(T)$ has arbitrarily many independent elements. Then Q, as a subgroup of $B_1(T)$, has an isomorphic image in $B_1(S)$, and so in $\prod B_1(G_n)$; and the argument in § 5 derives as before two mappings F, G of S in S^1 which correspond to elements of Q and which cannot be 2-compatible on S. But $F \mid T$ and $G \mid T$ are 2-compatible on T; hence (3) above is violated, and so $r(T)$ is not attained.

REFERENCES

1. S. Eilenberg, 'Sur les espaces multicohérents (I)', *Fund. Math.* 27 (1936), 153–90.
2. C. Kuratowski, *Topologie* (II) (Warsaw, 1950).
3, 4, 5. A. H. Stone, 'Incidence relations in multicoherent spaces (I, II, III)': (I) *Trans. American Math. Soc.* 65 (1949), 427–47; (II) *Canadian J. Math.* 2 (1950), 461–80; (III) *Pacific J. Math.* 2 (1952), 99–126.
6. G. T. Whyburn, *Analytic topology* (New York, 1942).

PATH GEOMETRY AND CONTINUOUS GROUPS

By D. D. KOSAMBI (*Bombay*)

[Received 31 March 1952]

THIS note attempts to show the relationship between topological groups and the differential geometry of path-spaces.

1. The paths of a path-space are defined as the real solutions in $R^n(x^1,\ldots,x^n)$ of the system of ordinary differential equations:

$$\ddot{x}^i+\alpha^i(x,\dot{x},t)=0 \qquad (\dot{x}^i=dx^i/dt,\ \ddot{x}^i=d\dot{x}^i/dt). \tag{1.1}$$

For a proper geometrical exploration, it would be necessary to have t an arc-like rather than an absolute time-like parameter; also, to have just one path through each direction at every regular point. The tangent vector $\dot{x}$ in Euclidean geometry is taken to define the same direction as $a\dot{x}$ for any constant non-zero factor a, which means allowing for a scale factor in t. The paths (1.1) must, therefore, have a restricted form which remains unchanged† for $\bar{t}=at+b$: that is,

$$\ddot{x}^i+\alpha^i(x,\dot{x})=0, \qquad \partial\alpha^i/\partial t=0, \qquad \epsilon^i\equiv\alpha^i-\tfrac{1}{2}\dot{x}^r\alpha^i_{;r}=0; \tag{1.2}$$

a semicolon and a comma in subscript indicate partial differentiation in $\dot{x}$ and in x respectively. Besides requiring this restricted form of the path-equations, we further demand that through any two sufficiently near points of the base-space there should pass one and only one path. This means assuming sufficient differentiability and some Lipschitz conditions;‡ for any x, the α^i are not much more general than quadratic homogeneous polynomials in $\dot{x}$.

Definition. A point of R^n belongs to (the manifold or base-space of) the restricted path-space of equations (1.2) if through the point there passes just one solution in each direction and if any two such points with $\max|x-x'|<\epsilon$ are connected by one and only one such path, for ϵ sufficiently small.

The manifold possesses a norm, say $\max|x-x'|$, and a corresponding topology that has no direct relation to the paths. We proceed to endow it with a special topology as follows. The arcs are paths or continuous

† This quite obvious result was first given by J. Douglas, *Annals of Math.* 29 (1928), 143–68.

‡ P. Hartmann and A. Wintner, *American J. Math.* 73 (1951), 149–72, and other references given there. The methods extend to our case.

curves made up of a finite number of path-segments. The new neighbourhood of a point P consists of all points that can be reached by paths emanating from P such that any two points of the neighbourhood, including P, are joined by one and only one path. Starting the paths from P at $t = 0$, a neighbourhood may then be defined by $t < T$. We consider only connected portions; any point that can be reached from an interior P by an arc, but which has not a complete path-bundle attached to it, belongs to the boundary.

A fixed point, strictly speaking, would also be a path, for $x = x_0, \dot{x} = 0$ satisfies (1.2). We exclude these by not permitting all initial $\dot{x}^i$ to vanish simultaneously at P, by some condition like $\sum (\dot{x}_0^i)^2 = 1$ (which will later be replaced by $g_{ij}\dot{x}_0^i\dot{x}_0^j = 1$). Then the coordinates of every point in the neighbourhood of a fixed P are determined with reference to those of P by means of t and $n-1$ of the initial direction-components, being in fact single-valued differentiable functions of n parameters; the parameters may, conversely, be expressed as such functions of the local coordinates. For the lens-space common to the two overlapping neighbourhoods of two sufficiently near points we have two different coordinate systems; the change from one to the other is described by differentiable functions of the coordinates themselves in either system. Therefore, by a classical result due to Whitney,† we have proved

THEOREM 1.1. *The base-space of any connected restricted path-space is a differentiable manifold, and may therefore be endowed with a (positive-definite) Riemann metric.*

This existence theorem gives no direct relationship between the path-equations (1.2) and the metric; the equations themselves need not (as they stand) be the extremals‡ of any variational principle at all, let alone geodesics of a particular metric. However, some indication of a 'best' choice and a special procedure is given by

THEOREM 1.2. *The neighbourhood of a point of any restricted path-space may be regarded as a neighbourhood of the identity in a topological group.*

Proof. We identify the fixed point O with the identity of the group. A group product law for two points P, Q in the neighbourhood is defined by taking QP as the point obtained on the path starting at $t = 0$ through P, in the same direction at P as for the path OQ at O; the value of t is to be taken as that needed to reach Q starting at $t = 0$ from O. This,

† H. Whitney, 'Differentiable manifolds', *Annals of Math.* 37 (1936), 645–80.

‡ The equations of variation of the path-equations must be self-adjoint; cf. D. D. Kosambi, *Quart. J. of Math.* (Oxford), 6 (1935), 1–12.

naturally, implies fixing of the parameter t in advance over the whole neighbourhood. The inverse of P is seen to be obtained by going backwards on OP through the origin, i.e. replacing t by $-t$. The associative law holds, for the product depends only upon the order of the points, the neighbourhoods being chosen so small that any pair of points in the locality determines the path-segment joining them.

A global extension of the local correspondence depends upon the original specification of the parameter, and the existence of a suitable O for the identity. Compactness or some such assumption simplifies the boundary-point question; but more than one point of the space may represent the same element of the whole group.

2. The converse is more fruitful.

THEOREM 2.1. *The connected component about the identity of any locally Euclidean topological group with suitably differentiable underlying manifold is a restricted path-space.*

Proof. An actual process of realization may be set up thus: with the identity as origin O, we choose a (local) system of coordinates which maps some neighbourhood of the origin into a corresponding neighbourhood of Euclidean space, E_n. Close to O, take a point P and generate a curve segment by multiplying with P (under the group law, say always from the right) the group-image of every point of the straight line segment in E_n joining OP. This is extended by operation upon the curve segment with its new end-point P^2, and so on. Letting P approach O along the fixed line segment, we obtain the path as the limiting curve through the origin for the fixed direction. This corresponds to a one-parameter semigroup, the rest of the group being obtained by operating with respect to P^{-1}, thus continuing the path through the origin.

Every one-parameter group is Abelian and, with suitable choice of the parameter, a translation (because of the exponential mapping) in that parameter. Assuming that this has been done, differentiability implies that the path has a tangent, whence the continuation through the origin is smooth at O, i.e. in the same direction but opposite sense.

The bundle of paths through O is now transferred to every point Q of the group manifold (we shall consider only the connected portion, but this causes no loss of generality) by taking the product with Q, always in some order fixed in advance. The identity of all the one-parameter groups (paths through O) may simultaneously be placed at $t = 0$; but there still remains an arbitrary scale-factor along each path, which can vary with the direction. However, the translation parameter may be

fixed along one chosen path, and the principle of continuity applied to restrict the others, t being required to vary continuously as P varies in any manner. Moreover, the scale factor is the same for any half-path and its continuation through O. These t values are then transferred from the neighbourhood of O to that of Q, along with the path-bundle, by the group-product. The general parameter t thus restricted still admits transformation, without changing the loci that give the paths; in particular, it always admits the general transformation $\bar{t} = at+b$ $(a \neq 0)$. Sufficient differentiability, assumed by hypothesis, means obeying second-order differential equations, which can always be solved for the highest derivatives, so that the paths may be described by equations of the restricted form (1.2). This completes the proof.

The process of standardizing the parameter t may be continued still further to obtain a Riemann metric for the group-manifold. The components of the metric tensor are *calculated at* O by identifying the parameter t *in the limit* with the distance s in each direction. Then this limiting process (and therewith the tensor g_{ij}) is transferred to each point of the manifold by the group product. This makes the *new* parameter t, now completely determined at every point along every direction and by integration along every arc, the arc-length in terms of the Riemannian metric tensor g_{ij}. We observe that the triangular inequality holds for the shortest distance between two points in terms of the Riemann metric; it holds therefore with small arcs consisting of paths. Now any simple curve joining OP, P near O, may be approximated by arcs made of joined path-segments, whence the path OP cannot be longer than any other curve. This gives us

THEOREM 2.2. *A Riemannian metric tensor g_{ij} may be specified for the connected portion about the identity of the group manifold, the paths then becoming the geodesics of this metric.*

For portions not connected, one cannot speak of 'distance' along an arc between any two points, but there is no other difficulty so long as some connected region exists about O. The essential process is that of determining t as a path-parameter over the entire manifold, and this is not necessarily unique, even apart from the fact that for a non-Abelian group we obtain two different sets of paths according to the order of operation.

The mechanism of transference from O to any other point of the manifold depends essentially upon the fact that *the connected topological group is also a simply transitive group of its own collineations*; hence of its

path-space. For a one-parameter subgroup, we should have a suitable generating vector $u^i(x)$ and the expansion

$$\bar{x}^i = x_0^i + \tau u^i(x_0) + \frac{\tau^2}{2!} X u^i + \ldots \quad (X = u^r \partial/\partial x^r). \tag{2.1}$$

On the other hand we have another formal expansion along any path:

$$\bar{x}^i = x_0^i + t\dot{x}_0^i + \frac{t^2}{2!}\alpha^i(x_0, \dot{x}_0) + \ldots . \tag{2.2}$$

For *any* tensor $T_{ij}^{\cdots}$, (2.1) leads to the formal expansion†

$$\left.\begin{aligned} \bar{T}_{ij}^{\cdots} &= T_{ij}^{\cdots} + \tau L T_{ij}^{\cdots} + \frac{\tau^2}{2!} L^2 T_{ij}^{\cdots} + \ldots \\ L\bar{T}_{ij}^{\cdots} &= u^r T_{ij,r}^{\cdots} + T_{ir}^{\cdots} u^r{}_{,j} + T_{rj}^{\cdots} u^r{}_{,i} + \ldots \end{aligned}\right\}. \tag{2.3}$$

Our derivation compels us to equate the parameters t in (2.2) and τ in (2.1) to the arc distance, provided that the generating vector u^i has been suitably chosen. This distance being preserved by transference, we have $Lg_{ij} = 0$ (the equations of Killing). This does *not* mean that the numerical value of the components g_{ij} are the same at every point, but that the components at any given point are determined in terms of those at the origin by the relationship between the coordinate systems at the two points $\overline{g_{ij}} = g_{rs} F_i^r F_j^s$, where the coefficients F_j^i equal $\partial x^i/\partial \bar{x}^j$, and also $\partial \dot{x}^i/\partial \dot{\bar{x}}^j$. This proves

THEOREM 2.3. *The relationship between the coordinate system at any point and that at O completely determines the components of g_{ij} at that point in terms of those at O, by the tensor law under change of coordinates.*

In particular, the coordinate system at O can always be taken as Euclidean, by suitable linear transformation, if necessary, to give $g_{ij}(0)$ the correct form. The theorem then reflects again Whitney's‡ result for any differential manifold, where g_{ij} is determined by the relationship between the local coordinate system mapping the neighbourhood into the Euclidean hypercube.

The differential geometry of the group, which, because of differentiability, has a Lie algebra of infinitesimal collineations defined by the vector solutions $u^i(x)$ of the equations of variations of the paths, can be determined in terms of the Lie constants of structure thus:

† D. D. Kosambi: *Proc. Nat. Acad. Sci.* (U.S.A.) **35** (1949), 389–94; *Quart. J. of Math.* (Oxford) (2) **2** (1951), 244–57; **3** (1952), 1–11. K. Yano: *J. Fac. Sci. Tokyo*, **6** (1949), 1–75; *Groups of Transformations in Generalized Spaces* (Tokyo, 1949).

‡ Loc. cit.

THEOREM 2.4. *The mixed Riemann–Christoffel curvature tensor R^i_{jkl} is given at O by $e^r_{jk}e^i_{rl}$, where the e's are the Lie constants of structure of the group; the determination of the curvature tensor elsewhere is by transference, $LR^i_{jkl}=0$.*

Proof. The final portion of the theorem follows from $Lg_{ij}=0$; or from the fact that the paths are preserved, and the tensor may be calculated directly from the path-equations. For the first part, the definition of the constants of structure is by

$$u^r_\alpha u^i_{\beta,r}-u^r_\beta u^i_{\alpha,r}\equiv L_\alpha u^i_\beta = e^\rho_{\alpha\beta}u^i_\rho. \tag{2.4}$$

The u^i_ρ are the n generating vectors of the group, whence $\det|u^i_\nu|$ cannot vanish; L_ρ is the corresponding Lie operator. For our special choice of parameter, at O, $L_\alpha=\partial/\partial s^\alpha$, the directional derivative in the αth direction at O. If we normalize the vectors at the origin, making the u^i_α into direction cosines, this operator would be $u^i_\alpha\nabla_i$, where ∇_i is the covariant derivative. Finally, the αth coordinate direction may be taken along u^i_α, which then has at O the value δ^i_α. At this stage, alternating $\partial^2/\partial s^\alpha\partial s^\beta$ at O amounts to alternating $L_\alpha L_\beta$ and also to alternating two covariant derivatives. The result of the latter operation upon any vector λ^i is

$$(\nabla_j\nabla_k-\nabla_k\nabla_j)\lambda^i = R^i_{jk\sigma}\lambda^r. \tag{2.5}$$

For the vector λ^i, we now choose one of the group generators, say u^i_γ. With (2.4) applied twice and $u^\xi_\nu=\delta^i_\nu$ in the final result, the theorem follows at once.

The properties of the curvature tensor

$$R^i_{jkl}+R^i_{kjl}=0, \qquad R^i_{jkl}+R^i_{klj}+R^i_{ljk}=0 \tag{2.6}$$

follow immediately from those of the Lie constants, namely

$$e^\alpha_{\beta\gamma}+e^\alpha_{\gamma\beta}=0, \qquad e^\rho_{\alpha\beta}e^\nu_{\rho\gamma}+e^\rho_{\beta\gamma}e^\nu_{\rho\alpha}+e^\rho_{\gamma\alpha}e^\nu_{\rho\beta}=0. \tag{2.7}$$

We further obtain the following consequences of (2.7),

$$R^r_{jkr}=0, \qquad R^r_{rkl}=R^r_{rlk}, \tag{2.8}$$

which are necessary for the existence of a metric. Yet the choice of a particular coordinate system at O entails one relation between the metric tensor at O and the constants $e^\alpha_{\beta\gamma}$, from

$$g_{ir}R^r_{klj}+g_{rj}R^r_{kli}=0, \tag{2.9}$$

which is then transferred to all points of the space; this is an identity for every Riemann space as well as the integrability condition† of the

† L. P. Eisenhart, *Non-Riemannian Geometry* (American Math. Soc. Colloquium Pub. VIII, New York 1927), § 29.

partial differential equations $g_{ij|k} = 0$, i.e. the vanishing of the covariant derivative for the metric tensor. The group generators u_σ^i may always be replaced by linear combinations with constant coefficients, which then means transforming the $e_{\beta\gamma}^\alpha$ like the components of a tensor in the Greek indices.

3. The results of the previous section determine the differential geometry of a given differentiable group manifold. Some further steps may be taken, as for example identifying $e_{\beta\gamma}^\alpha$ with the components of the Cartan *torsion* tensor at the origin; in that case, the Riemannian curvature is determined by the torsion.† Moreover, it can be shown‡ that an *antisymmetric component can be assigned to* g_{ij} *whose exterior derivative is a constant multiple of the fully covariant torsion tensor.* The total curvature tensor vanishes when such torsion is permitted, so that we have spaces with teleparallelism; but it simplifies the geometry to take the space with zero torsion though with non-vanishing curvature.

The metric allows overall invariants of the manifold and therefore of the group to be calculated by known methods,§ such as extensions of the Gauss–Bonnet formula. If the group is a direct product, the manifold will necessarily be a product space, with a corresponding reduction of the metric into several portions each involving only such coordinates and their differentials as do not enter into any other portion.

Though the group determines a Riemann space, the Riemann space does *not* necessarily determine the group uniquely. For example

The manifold of an Abelian group is flat, so that its metric may be taken as Euclidean, and conversely.

The direct result is obvious, but the converse is more complicated algebraically. If the curvature tensor vanishes, we have only $e_{\beta\gamma}^\gamma e_{\rho\alpha}^\nu = 0$, which does not immediately lead to $e_{\alpha\beta}^\rho = 0$. At first, we need only have $e_{\alpha\beta}^\rho = 0$ for all α, β, if $\rho > m$; also for all ρ with $\alpha, \beta \leqslant m$ where $m \leqslant n$, so that we have an m-dimensional subgroup all of whose operators commute with those of the entire group; and m cannot be zero. Taking

† The first step seems to have been taken by L. P. Eisenhart, but the real development is due to Cartan and Schouten, cf. L. P. Eisenhart, *Continuous Groups of Transformations* (Princeton 1933), 192–207; bibliography, 294–5.

‡ D. D. Kosambi, *Bull. American Math. Soc.* 55 (1949), 90–4.

§ That is, for closed Riemannian manifolds: C. B. Allendoerfer and A. Weil, *Trans. American Math. Soc.* 53 (1943), 101–29; S. S. Chern, *Annals of Math.* 45 (1944), 747–52; *Proc. Nat. Acad. Sci.* (U.S.A.) 30 (1944), 269–73; *Annals of Math.* 46 (1945), 674–84. S. Bochner, ibid. 49 (1948), 379–90 and 50 (1949), 77–93.

this m-dimensional Abelian subgroup as translations in the first m coordinates, we see that the rest cannot be made up of rotations. Two infinitesimal rotations in the coordinate directions (ij), (jk) combine to give (ik), so that there are not enough disjoint rotations and, if there are at least two such, there must be a third in the group, and one component of the curvature tensor will not vanish. Moreover, these cannot contain any of the translation coordinates, because that subgroup permutes with all operators of the n-group. Thus the entire n-group consists of translations. The argument could have been shortened by recalling that we must have a simply transitive n-parameter subgroup of Euclidean rigid motions. Another simplification would be to note that the flat space is its own tangent space; then G. Birkhoff's theorem applies,† to the effect that the original space is the homogeneous space G/S, where G is the whole group of isomorphisms, S the (local) rotation group.

Starting with a group and its manifold, the example shows us that there exist cases where the group itself is a proper subgroup of its own group of automorphisms. For example, the manifold of the Abelian group, as a Euclidean space, not only admits the group itself, *translations*, but also, in the coordinate system chosen, a group of rotations preserving the metric; while the path-space admits the general linear group, preserving the paths (straight lines) though not the metric. We have really a special case of

THEOREM 3.1. *If the constants of structure obey*

$$e^{\rho}_{\alpha\beta}\, e^{\nu}_{\rho\gamma} = \frac{1}{n-1}(\delta^{\nu}_{\alpha} R^{0}_{\beta\gamma} - \delta^{\nu}_{\beta} R^{0}_{\alpha\gamma}), \qquad R^{0}_{\beta\gamma} = e^{\rho}_{\nu\beta}\, e^{\nu}_{\rho\gamma}, \tag{3.1}$$

the group-manifold is an isotropic Riemann space; hence the group itself is an n-dimensional subgroup of an $\frac{1}{2}n(n+1)$-dimensional subgroup of the full projective group in the n coordinate variables.

Proof. The relations (3.1) and transference make the Riemann manifold isotropic (projectively flat, of zero Schur constant). It is a well-known result that such a manifold admits locally a group of dimension $\frac{1}{2}n(n+1)$, and the fact that we have the whole group here restricts the group and the space to a few classical types,† all of which are included in the general statement of the theorem. However, in view of the fact that path-geometry is wider than that of the group-manifold, it may be as well to extend the concept of isotropy and the result of the theorem to paths of type (1.1) with only $\partial\alpha^{i}/\partial t = 0$. This is done as follows.

† Garrett Birkhoff, *Math. Zeits.* **53** (1950), 226–35.

When there is a metric, a subspace has paths which can be derived as the extremals, in the coordinates y, which define the subspace by the equations $x^i = x^i(y^\alpha)$ $(i = 1, 2,..., n;\ \alpha = 1, 2,..., m)$.

Without a metric, this cannot be done uniquely. But suppose that some system of paths

$$\ddot{y}^\nu + \beta^\nu(y, \dot{y}) = 0$$

be given *a priori* in the subspace, with the additional property that *every path of the subspace is a path of the original space*, the subspace being thus a *totally geodesic* variety.

THEOREM 3.2. *For a totally geodesic subspace, the path-functions obey the relationship*

$$x^i{}_{,\rho,\sigma}\dot{y}^\rho\dot{y}^\sigma - x^i_\rho\beta^\rho + \alpha^i\{x(y), x^j{}_{,\nu}\dot{y}^\nu\} = 0. \tag{3.2}$$

Moreover, the corresponding differential invariants of the original space and the subspace are related† *by*

$$T^i_j : x^j{}_{,\alpha} + \ldots = T^\nu_\alpha . x^i{}_{,\nu} + \ldots . \tag{3.3}$$

The proof is by direct substitution and elimination.

Definition. A space is *projectively flat* if it admits (locally) a system of coordinates such that every subspace determined by $y^i = x^i$ $(i = 1,..., m)$ and $x^i = 0$ $(i = m+1,..., n)$ is totally geodesic.

THEOREM 3.3. *For a projectively flat space, the paths have the equations*

$$\ddot{x}^i + \dot{x}^i\phi(x, \dot{x}) = 0. \tag{3.4}$$

This again follows by direct substitution of the special coordinate values in the above relationship.

The equations of variation define the collineations, i.e. the automorphism group preserving the paths, which is given by the vector generators $u^i(x)$ that satisfy

$$u^i{}_{,j,k}\dot{x}^j\dot{x}^k + \dot{x}^i L\phi = 0, \quad \text{where } L\phi \equiv u^r\phi_{,r} + u^r{}_{,s}\dot{x}^s\phi_{,r}. \tag{3.5}$$

It follows that $L\phi$ can at most have the form $\dot{x}^r a_r(x)$. Now the operator $\partial/\partial\dot{x}^j$ permutes with L, and L is a differentiation for which $L\dot{x}^i = 0$, regardless of the group generator employed. By differentiation of (3.5) twice in $\dot{x}$, we have therefore

$$(a)\ \ 2u^i{}_{,j,k} + \delta^i_j a_k + \delta^i_k a_j = 0,$$

$$(b)\ \ -2(u^i{}_{,j,k,l} - u^i{}_{,j,l,k}) \equiv \delta^i_j(a_{k,l} - a_{l,k}) + \delta^i_k a_{j,l} - \delta^i_l a_{j,k} = 0. \tag{3.6}$$

† The quickest method of proving this statement is to calculate the differential invariants of (3.2) by regarding those equations as defining a path-geometry of a special type; cf. D. D. Kosambi, *Quart. J. of Math.* (Oxford), 19 (1948), 204–19.

In the second, we contract successively for $i-j$ and $i-k$, to obtain $a_{i,j}=0$, and also $u^i{}_{,j,k,l}=0$. These results are summed up in

THEOREM 3.4. *The group generators of a projectively flat space with paths* $\ddot{x}^i+\dot{x}^i\phi(x,\dot{x})=0$ *have the form* $u^i=-\frac{1}{2}x^i a_r x^r+b^i_r x^r+c^i$ *where the constants* a_i *must obey the further restriction* $L\phi_{,r}=a_r$.

This proves our statement† that the group of all automorphisms is a *subgroup of the projective group in n variables.* The full projective group is not admissible, for it has $n(n+2)$ parameters, whereas the maximum number possible for any path-space (without reference to a metric) is only $n(n+1)$, and this number is realized if and only if the space is flat, the paths then being straight lines and the group that of translations plus the full linear homogeneous group. When a Riemann metric is to be preserved, the parameters reduce to $\frac{1}{2}n(n+1)$ in number, being the maximum admitted by any such metric.

Regardless of the space being a topological group or not, we can derive conditions of isotropy for an affinely connected manifold. Here, $\phi(x,\dot{x})=\phi_r\dot{x}^r$, the vector ϕ_i depending only upon the coordinates x, not upon the directions. The coefficients of connexion are

$$\Gamma^i_{jk}=\tfrac{1}{2}(\delta^i_j\phi_k+\delta^i_k\phi_j). \tag{3.7}$$

Calculating the curvature tensor defined by

$$R^i_{jkl}=\Gamma^i_{lj,k}-\Gamma^i_{lk,j}+\Gamma^i_{rk}\Gamma^r_{lj}-\Gamma^i_{rj}\Gamma^r_{lk}, \tag{3.8}$$

we find its components to be

$$R^i_{jkl}=\tfrac{1}{2}\{\delta^i_j(\phi_{l,k}-\tfrac{1}{2}\phi_k\phi_l)-\delta^i_k(\phi_{l,j}-\tfrac{1}{2}\phi_j\phi_l)+\delta^i_l(\phi_{j,k}-\phi_{k,j})\}. \tag{3.9}$$

When a Riemann metric exists, we further obtain:

(1) The vector ϕ_i is a gradient $\phi_{,i}$, for $\phi_{i,j}-\phi_{j,i}=0$, from $R^r_{jkr}=0$.

(2) Either the space is flat and the curvature tensor vanishes identically, in which case $\phi=-\log(a_r x^r+b)^2$; or, the curvature constant K not being zero, we have the canonical form $g_{ij}=(\phi_{i,j}-\frac{1}{2}\phi_{,i}\phi_j)/2K$.

(3) From the vanishing of the covariant derivative of g_{ij}, reducing to $\{\exp(-\phi)\}_{i,j,k}=0$, we see when $K\neq 0$ that $\phi=-\log(a_{ij}x^ix^j+2b_ix^i+c)$.

(4) A further reduction follows from $g_{ij}\neq 0$, with

$$\left.\begin{aligned} g_{ij}&=-\frac{1}{K}\left\{\frac{a_{ij}}{S}-\frac{(a_{ir}x^r+b_i)(a_{js}x^s+b_j)}{S^2}\right\}\\ S&=a_{ij}x^ix^j+2b_ix^i+c\end{aligned}\right\}. \tag{3.10}$$

† For $n=2$, the general, non-parametric path-space can be shown, by direct calculation of all the groups, to admit only subgroups of the projective group; cf. D. D. Kosambi, *Quart. J. of Math.* (Oxford), (2) **3** (1952), 1–11.

A linear change of coordinates x^i does not affect the special form of the paths. It follows (by suitable transformation) that S must be non-degenerate, for otherwise there would be a singularity at the origin; S must be definite, otherwise ds^2 would change sign at O for different directions.

THEOREM 3.5. *The group of automorphisms of an n-dimensional isotropic Riemann space is either that of translations and rotations (flat space), or equivalent to the projective group of a non-degenerate central quadric.*

It follows as a corollary that *the only groups that are isotropic spaces* ($n > 3$) *are Abelian, the spaces being flat.*

For, in the non-flat case, we may reduce the quadric to $\sum (x^i)^2 - 1 = 0$, by an always-permissible linear transformation. The generators of the group G constitute the Lie algebra of operators $x^i p_j - x^j p_i$ of rotation S, and the projections $p_i - x^i x^r p_r$. The last do not form an ideal or a sub-algebra of the Lie algebra, the commutator of any two giving a rotation operator. Therefore the homogeneous space G/S cannot be a group; yet it is isomorphic to the original space.

A theorem of Cartan and Schouten,† that the space of a simply transitive semi-simple group is an Einstein space, with g_{ij} proportional to R_{ij}, includes the case of groups that are isotropic spaces.

Every path-space (1.1) can be reduced to the restricted form (1.2) by taking $x^0 = t$ as an extra coordinate and a new implicit parameter; but the process is not unique, while the varieties $x^0 =$ constant have no meaning for the original space, corresponding to impulsive motion in dynamics, just as the restricted path-spaces give *kinematic* trajectories. On the other hand, we get a rather obvious inverse result that justifies the name 'projectively flat':

THEOREM 3.6. *The paths of a restricted projectively flat path-space reduce to straight lines* $d^2x^i/d(x^n)^2 = 0$ *by taking* x^n *as a new parameter in place of t.*

[*Added* 8 *May* 1952]

4. The foregoing gives some insight into the mechanism of locally Euclidean topological groups with (orientable) differentiable manifolds. Its essence has now to be recapitulated for path-spaces that are not assumed to be groups, for there may be some difference between the topological group and its representation as a path space; this is seen from

† See the first footnote on p. 313.

the fact that a *compact* Abelian group is topologically a torus, which cannot be a space of zero curvature. The particular difficulty, of course, is resolved by taking a hypercube whose opposite faces are identified.

For paths $\ddot{x}^i+\alpha^i(x,\dot{x},t)=0$, there is always the parameter of total differentiation d/dt, with a corresponding exponential mapping:

$$\left.\begin{aligned} &(a)\quad \frac{d}{dt}\equiv\frac{\partial}{\partial t}+\dot{x}^r\frac{\partial}{\partial x^r}-\alpha^r\frac{\partial}{\partial\dot{x}^r}\\ &(b)\quad \overline{F}(x,\dot{x},t)=F+t\frac{dF}{dt}+\frac{t^2}{2!}\frac{d}{dt}\left(\frac{dF}{dt}\right)+\dots\\ &(c)\quad x^i=ta^i-\frac{t^2}{2!}\alpha^i_0-\frac{t^3}{3!}\left(\frac{d\alpha^i}{dt}\right)_0+\dots \end{aligned}\right\}. \tag{4.1}$$

In the last formal expansion the point x^i is reached by starting from the origin (chosen arbitrarily) in direction $\dot{x}^i_0=a^i$, $\sum(a^i)^2=1$, to get a locally Euclidean correspondence for some neighbourhood of O, and hence for the whole path-space. To permit this, we must first assume that the formal series either converges, or at least represents in some unique fashion the point x^i in (4.1 *c*). The method, of course, is to build up a sum of successive small translations in the parameter, say by overlapping intervals, which give our formalism its meaning. Secondly, $x^i(ta)$ must be solvable to give $ta^i=f(x)$, the ta^i being then coordinates that map a sufficiently small neighbourhood of O into that of E_n. This means the vanishing of the two differential invariants $\alpha^i-\frac{1}{2}\dot{x}^r\gamma^i_{;r}$ and $\partial\alpha^i/\partial t$.

The coefficients $F^i_j=\partial\bar{x}^i/\partial x_j$ are obtained for the local correspondence $0\longleftrightarrow x$ by $\dot{x}^i=F^i_ra^r$. From differentiation of (4.1 *c*) and the homogeneity of α^i, we factor the formal series

$$\dot{x}^i=a^i-t\alpha^i_0-\frac{t^2}{2!}\left(\frac{d\alpha^i}{dt}\right)_0\dots=a^r\{\delta^i_o+H^i_r\}=a^rF^i_r. \tag{4.2}$$

From inverting (4.1 *c*) for ta^i in terms of x^i, we obtain $F^i_j=F^i_j(x)$; it is seen that the homogeneity of α^i permits the simultaneous elimination of ta^i in (4.2) after a^r has been factored out just once. The determinant $|F^i_j|$ being unity at O will remain positive for t sufficiently small, which makes the neighbourhood orientable.

In this $0\longleftrightarrow x$ correspondence, the tangent vector $\dot{x}^i$ to the path-segment (necessarily assumed unique) joining O to x always corresponds to the initial direction a^i of the path. Since O was arbitrary and we can always go as far as desired by overlapping neighbourhoods, this means geometrically that the *paths are autoparallel lines*.

As seen from the treatment by group theory, a change of parameter will in general be necessary to reduce the path-equations explicitly to a suitable form. This change will depend not only upon the point but also upon the direction; s being another parameter, $ds/dt = \psi(x, \dot{x})$, with ψ homogeneous of degree one in $\dot{x}$. After such a change, α^i with the new parameter must reduce to those of a Riemann space, hence in the first instance to quadratic polynomials in $\dot{x}$. Direct calculation shows that this amounts to replacing α^i by $\alpha^i + \dot{x}^i\phi(x, \dot{x})$ (with a corresponding recalculation of the formal relation between coordinates $0 \longleftrightarrow x$), where ϕ is still homogeneous of the first degree in $\dot{x}$. This is called a *general projective change of parameter*; such change must make the path-space a Riemann space. The only essential differential invariant of the path-space which is independent of such change is the generalized Weyl tensor

$$W^i_{jk} = R^i_{jk} - \dot{x}^i R^r_{jk;r} + \frac{1}{n-1}\delta^i_j(R^r_{kr} - \dot{x}^s R^r_{ks;r}) - \frac{1}{n-1}\delta^i_k(R^r_{jr} - \dot{x}^s R^r_{js;r}), \tag{4.3}$$

where

$$R^i_{jk} = \tfrac{1}{3}(P^i_{j;k} - P^i_{k;j}), \qquad P^i_j = -\alpha^i{}_{,j} + \frac{1}{2}\frac{d}{dt}\alpha^i{}_{;j} + \tfrac{1}{4}\alpha^i{}_{;r}\alpha^r{}_{;j}.$$

THEOREM 4.1. *A path-space with unique path-segments joining sufficiently near points admits a Riemann metric by projective change of parameter, and is completely characterized by the necessary and sufficient conditions* $\partial\alpha^i/\partial t = 0$, $\alpha^i - \frac{1}{2}\dot{x}^r\alpha^i{}_{;r} = 0$, $W^i_{jk;l;m} = 0$ *(given also sufficient differentiability).*

The proof is obvious from the foregoing, because W^i_{jk} must be linear in $\dot{x}$, as for a Riemann space. We obtain a simple corollary, that *the paths of any symmetric affine connexion may be made the geodesics of a Riemann space by suitable projective choice of parameter, provided that the paths have the property of uniqueness between two sufficiently near points.*

For this corollary, we can give a direct construction as proof. The contraction $R^r_{jk;r}$ is always a curl and may be made to vanish by suitable choice of the vector μ_i in adding $\frac{1}{2}(\delta^i_j\mu_k + \delta^i_k\mu_j)$ to Γ^i_{jk}, which is the effect of the projective change of parameter. We assume this to have been done, the vector μ_i being then determined to within an additive gradient. This arbitrariness is used, if necessary, in making

$$g_{ir}\bar{R}^r_{kl;j} + g_{rj}\bar{R}^r_{kl;i} = 0$$

at the origin, where the tensor g_{ij} can always be taken as δ_{ij} at O. The general tensor g_{ij} is now determined by the recalculated $0 \longleftrightarrow x$

correspondence and the tensor law of transformation, which makes the covariant derivative of that tensor vanish; in fact the above relation with the curvature tensor will hold at every regular point if it holds at the origin, and that is the integrability condition for $g_{ij|k} = 0$. Finally, the determinant $|g_{ij}|$ will not vanish in general, since it was taken to be unity at O, and its partial derivative is a multiple of the contraction Γ^r_{ir}.

This seems to indicate that the projective curvature tensor, and not the Riemann–Christoffel tensor, should play the fundamental role in calculating overall invariants of path-spaces and group-spaces.

[*Added* 19 *May* 1952]

The anchor-ring in E_3 is the closed compact S_2 of genus one

$$x^1 = (a+b\sin\theta)\cos\phi, \qquad x^2 = (a+b\sin\theta)\sin\phi, \qquad x^3 = b\cos\theta.$$

The metric being

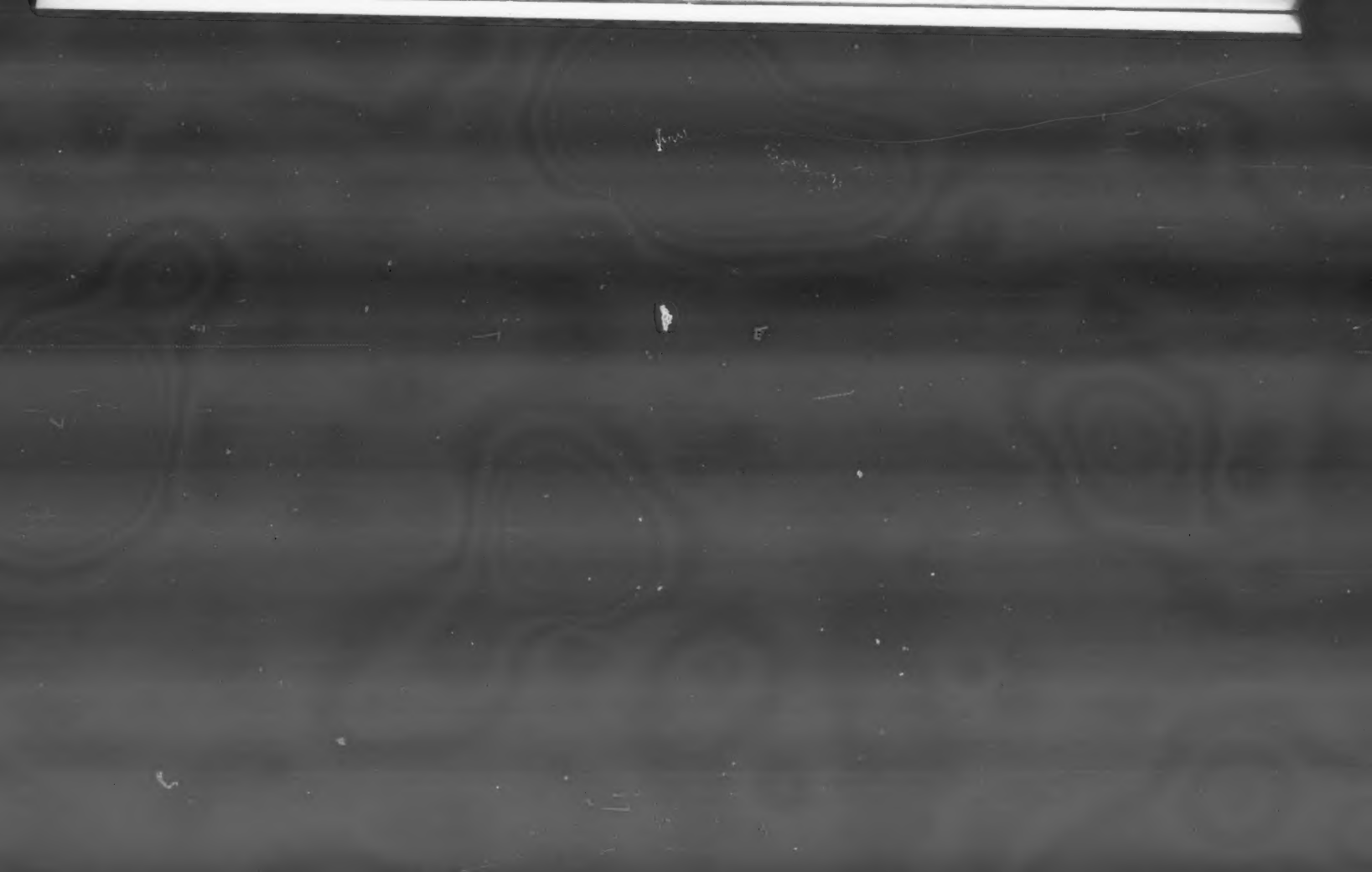
$$ds^2 = b^2\,d\theta^2+(a+b\sin\theta)^2\,d\phi^2,$$

no coordinate system exists making all the geodesics simultaneously into straight lines. In E_4, however, the intersection of two cylinders

$$x^1 = a\cos\theta, \qquad x^2 = a\sin\theta, \qquad x^3 = b\cos\phi, \qquad x^4 = b\sin\phi$$

has the metric $a^2\,d\theta^2+b^2\,d\phi^2$ of zero Gaussian curvature. It is only in this second form or its equivalent that the torus can represent a compact Abelian G_2; not as the anchor-ring torus. In the large, the difference appears in the imbedding properties of the group-space. The argument extends obviously to higher dimensions.

The developments of the first two sections give us, starting from a locally Euclidean group with suitably restricted paths, a Lie algebra of the infinitesimal collineations of the path-space into itself. The structure of this Lie algebra being determined entirely by the paths, hence by the original group, we may use the algebra to define an analytic group and manifold. It follows that the original topological group-space can be mapped upon the analytic manifold, again identifying the one-parameter groups and their cosets with the original paths. This exposes the structure of the Hilbert conjecture (fifth problem) for locally Euclidean groups that are also path-spaces; the problem is of little interest for path-geometry.



Advanced Dynamics

Two volumes

E. HOWARD SMART, M.A.

40s. each

Advanced Five-Figure Mathematical Tables

C. ATTWOOD, B.Sc., B.Sc.(Eng.), A.M.I.E.E.

6s.

The Creation of the Universe

PROF. GEORGE GAMOW

Illustrated

21s.

Intermediate Geometry

L. J. LACY, B.Sc.

Macmillan and Co., Ltd.

HEFFER'S

of Cambridge

for

BOOKS ON

MATHEMATICS

New & Secondhand

English & Foreign

W. HEFFER & SONS LTD.

Petty Cury, Cambridge

Hutchinson's

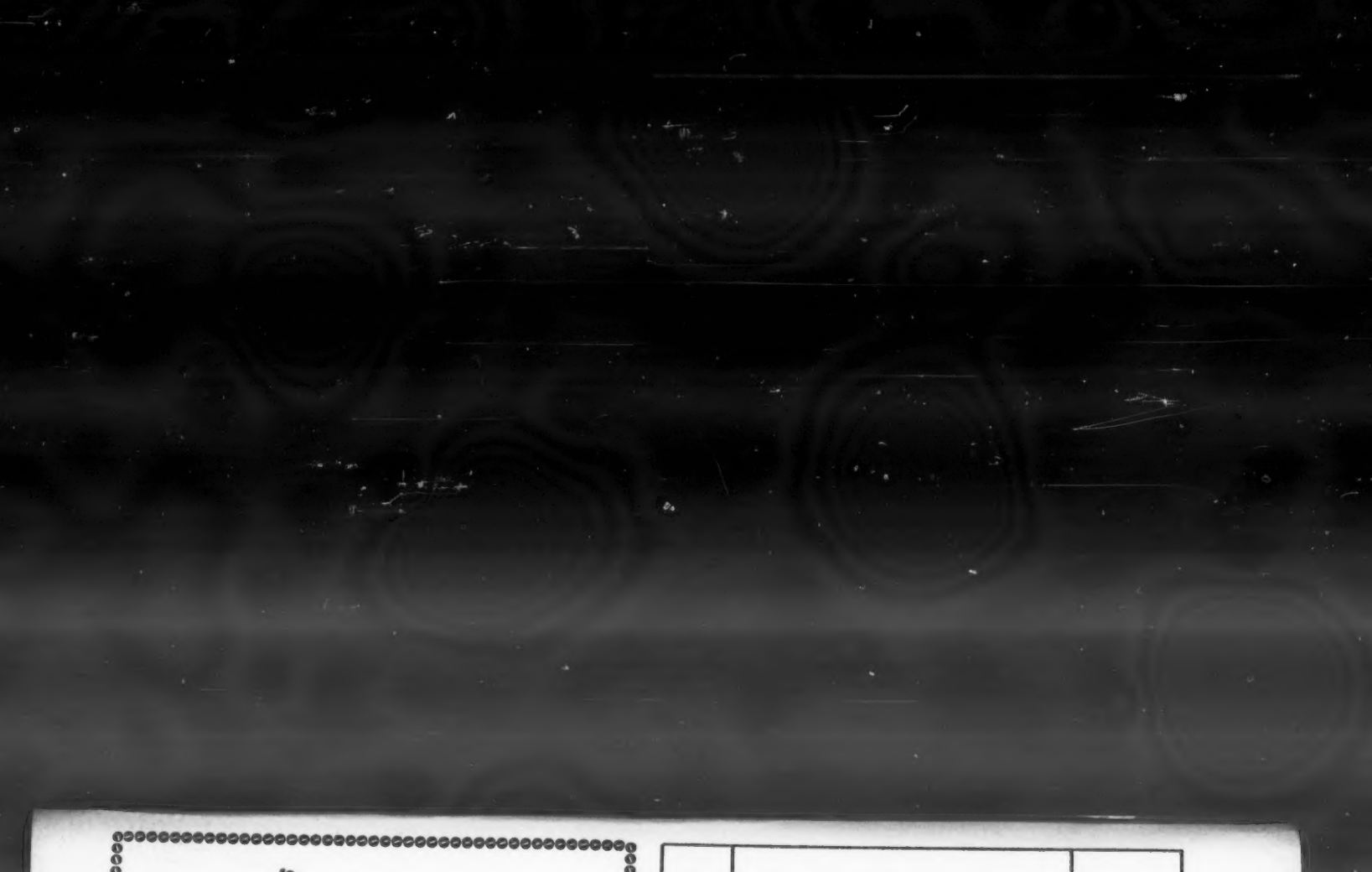
Just out:

THE HIGHER ARITHMETIC

Professor H. Davenport

Gives an introductory account of the theory of numbers, which embraces the properties of the natural numbers 1, 2, 3.

. . . .

8s. 6d. net

University Library